Forest Image Classification Based on Deep Learning and Ontologies

Clopas Kwenda (221072651)

A thesis submitted in fulfilment of the requirement for the degree of

Doctor of Philosophy in Computer Science

School of Mathematics, Statistics and Computer Science
University of KwaZulu-Natal,
Pietermaritzburg, South Africa

December, 2023

Forest Image Classification Based on Deep Learning and Ontologies

Clopas Kwenda (221072651)

Supervisor: Dr. Mandlenkosi Victor Gwetu

Co-Supervisor: Dr. Jean Vincent Fonou-Dombeu

A thesis submitted in fulfillment of the requirement for the degree of

Doctor of Philosophy in Computer Science

School of Mathematics, Statistics and Computer Science

University of KwaZulu-Natal

South Africa

EXAMINER'S COPY

December, 2023

As the candidate's supervisor, I have approved this thesis for submission.

Signed.....[redacted].....Date.....2 April 2024.....

Declaration 1 - Plagiarism

I, **Clopas Kwenda**, declare that;

1. The research reported in this thesis, except where otherwise indicated, is my original research.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
   (a) Their words have been re-written but the general information attributed to them has been referenced,
   (b) Where their exact words have been used, then their writing has been placed in italics and inside quotation marks and referenced.
5. This thesis does not contain text, graphics, or tables copied and pasted from the Internet, unless specifically acknowledged, and the source is detailed in the thesis and in the references sections.

Signed:...  ..... Date:02/04/2024.....

Declaration 2 - Publication

I, clopas Kwenda, declare that the following are publications from this thesis:

1. Kwenda, Clopas, Mandlenkosi Gwetu, and Jean Vincent Fonou-Dombeu. "Machine Learning Methods for Forest Image Analysis and Classification: A Survey of the State of the Art." IEEE Access, 10 (2022), pp. 45290-45316, 2023. DOI: 10.1109/ACCESS.2022.3170049

2. Clopas Kwenda, Mandlenkosi Victor Gwetu and Jean-Vincent Fonou-Dombeu. "A critical survey of GEOBIA methods for forest image detection and classification", Geocarto International, 38 (1), pp. 1-38, 2023.
   https://www.tandfonline.com/doi/full/10.1080/10106049.2023.2256302

3. Kwenda, Clopas, Mandlenkosi Gwetu, and Jean Vincent Fonou-Dombeu, "Hybrid Learning Model for Satellite Forest Image Segmentation." In: Rutkowski, L., Scherer, R., Korytkowski, M., Pedrycz, W., Tadeusiewicz, R., Zurada, J.M. (eds) Artificial Intelligence and Soft Computing. ICAISC 2023. Lecture Notes in Computer Science (LNCS), vol 14126. Springer, Cham. https:doi.org/10.1007/978-3-031-42508-0_4

4. Kwenda, Clopas, Mandlenkosi Victor Gwetu, and Jean Vincent Fonou-Dombeu. "Forest Image Classification Based on Deep Learning and XGBoost Algorithm." In: Mikyška, J., de Mulatier, C., Paszynski, M., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M. (eds) Computational Science – ICCS 2023. ICCS 2023. Lecture Notes in Computer Science (LNCS), vol 10476. Springer, Cham. https://doi.org/10.1007/978-3-031-36027-5_16

5. Kwenda, Clopas, Mandlenkosi Gwetu, and Jean Vincent Fonou-Dombeu. "Ontology with Deep Learning for Forest Image Classification." Applied Sciences, 13(8), pp. 1-21, 2023. https://doi.org/10.3390/app13085060

6. Kwenda, Clopas, Mandlenkosi Gwetu, and Jean-Vincent Fonou Dombeu."Hybridizing Deep Learning and Machine Learning Models for Aerial Satellite Forest Image Segmentation". Submitted to Applied Sciences. Under Review.

Dedication

This work is dedicated to the Almighty, the creator of heavens and earth

Acknowledgements

Firstly l want to issue my deepest thanks to the Almighty for providing good health, guidance, protection, and adequate financial provisions to undertake this study. I am indebted to the Great Zimbabwe University and the University of KwaZulu Natal for the financial assistance to cater for my fees and living subsistence.

I also want to extend my profound thanks to my supervisors Dr. Mandlenkosi Victor Gwetu and Dr Jean Vincent Fonou-Dombeu for their guidance and support throughout this study. l appreciate the constructive advice l received from my research colleagues and in particular: Emmerson Chivhenge and Wilson Bakasa.

I am sincerely grateful to my late father Genasio Kwenda and my mother Lena Kwenda. I appreciate your guidance and love.

Finally, I am indebted to my lovely wife Adelaide Ruzani Jirivengwa, and my children: Shannon and Shane. Your love and support is second to none. Thank you so much!

Abstract

Forests contribute abundantly to nature's natural resources and they significantly contribute to a wide range of environmental, socio-cultural, and economic benefits. Classifications of forest vegetation offer a practical method for categorising information about patterns of forest vegetation. This information is required to successfully plan for land use, map landscapes, and preserve natural habitats. Remote sensing technology has provided high spatio-temporal resolution images with many spectral bands that make conducting research in forestry easy. In that regard, artificial intelligence technologies assess forest damage. The field of remote sensing research is constantly adapting to leverage newly developed computational algorithms and increased computing power. Both the theory and the practice of remote sensing have significantly changed as a result of recent technological advancements, such as the creation of new sensors and improvements in data accessibility. Data-driven methods, including supervised classifiers (such as Random Forests) and deep learning classifiers, are gaining much importance in processing big earth observation data due to their accuracy in creating observable images. Though deep learning models produce satisfactory results, researchers find it difficult to understand how they make predictions because they are regarded as black-box in nature, owing to their complicated network structures. However, when inductive inference from data learning is taken into consideration, data-driven methods are less efficient in working with symbolic information. In data-driven techniques, the specialized knowledge that environmental scientists use to evaluate images obtained through remote sensing is typically disregarded. This limitation presents a significant obstacle for end users of Earth Observation applications who are accustomed to working with symbolic information, such as ecologists, agronomists, and other related professionals. This study advocates for the incorporation of ontologies in forest image classification owing to their ability in representing domain expert knowledge. The future of remote sensing science should be supported by knowledge representation techniques such as ontologies. The study presents a methodological framework that integrates deep learning techniques and ontologies with the aim of enhancing domain expert confidence as well as increasing the accuracy of forest image classification. In addressing this challenge, this study followed the

following systematic steps (i) A critical review of existing methods for forest image classification (ii) A critical analysis of appropriate methods for forest image classification (iii) Development of the state-of-the-art model for forest image segmentation (iv) Design of a hybrid model of deep learning and machine learning model for forest image classification (v) A state-of-the-art ontological framework for forest image classification. The ontological framework was flexible to capture the expression of the domain expert knowledge. The ontological state-of-the-art model performed well as it achieved a classification accuracy of 96%, with a Root Mean Square Error of 0.532. The model can also be used in the fruit industry and supermarkets to classify fruits into their respective categories. It can also be potentially used to classify trees with respect to their species. As a way of enhancing confidence in deep learning models by domain experts, the study recommended the adoption of explainable artificial intelligence (XAI) methods because they unpack the process by which deep learning models reach their decision. The study also recommended the adoption of high-resolution networks (HRNets) as an alternative to traditional deep learning models, because they can convert low-resolution representation to high-resolution and have efficient block structures developed according to new standards and they are excellent at being used for feature extraction.

Contents

List of Acronyms

ADI	Area Discrepancy Index
AIR	Actual Image Region
AOI	Area Of Interest
CE	Commission Error
CNN	Convolutional Neural Networks
CRF	Conditional Random Fields
DIR	Delineated Image Region
DL	Deep learning
DSM	Digital Surface Mode
FAO	Food and Agricultural Organization
GEOBIA	Geographic Object Based Image Analysis
GNB	Gaussian Naïve Bayes
GOFAI	Good Old Fashioned Artificial Intelligence
GOSE	Over Segmentation Error
GS	Global Score
GUSE	Under Segmentation Error
HIS	Hyper Spectral Image
IoU	Intersection Over Union
KNN	K Nearest Neighbor

LCCS	Land Cover Classification Systems
LDA	Linear Discriminant Analysis
LGBM	Light Gradient Boost Machine
LSMS	Large Scale Mean Shift
LSVM	Linear Support Vector Machine
LULC	Land Use and Land Cover
MAE	Mean Average Error
ML	Machine Learning
NDVI	Normalised Vegetation Vegetation Index
NPP	Net Primary Productivity
OBOE	Extensible Observation Ontology
OE	Ommission Error
OOB	Out Of Bag
OWL	Web Ontology Language
PCA	Principal Component Analysis
RE	Rand Error
RF	Random Forest
RMSE	Root Mean Square Error
ROC	Reciever Operating Characteristic
RS	Remote Sensing
SRC	Sparse Representation based Classification
SVM	Support Vector Machine
TL	Transfer Learning
TSNE	T Distributed Stochastic Neighbor Embedding
VHR	Very High Resolution
WT	Watershed Transform

XGBoost Extreme Gradient Boosting

Preface

The research presented in this thesis was carried out in the College of Agriculture, Engineering and Science of the University of Kwa-Zulu Natal, Durban, from January 2021 until December 2023 by Clopas Kwenda under the supervision of Dr. Mandlenkosi Victor Gwetu and co-supervised by Dr. Jean Vincent Fonou-Dombeu.

As the candidate's supervisor, I, Mandlenkosi Victor Gwetu, agree to the submission of this thesis.

[redacted]. [redacted]..... Date: 2 April 2024

As the candidate's co-supervisor, I, Jean Vincent Fonou-Dombeu, agree to the submission of this thesis.

Signed: [redacted]..... Date: 02 April 2024

I, Clopas Kwenda, hereby declare that all the material incorporated in this thesis are my own original work, except where acknowledgement is made by name or in the form of a reference. The work contained herein has not been submitted in any form for any degree or diploma to any other institution.

Signed: [redacted]..... Date: 02/04/2024

University of KwaZulu-Natal, April 2, 2024

1 Introduction

1.1 Introduction

Forests contribute abundantly to nature's natural resources and significantly contribute to a wide range of environmental, socio-cultural, and economic benefits [1]. While forests play a significant economic role in developing countries through the creation of jobs, income, and foreign exchange earnings, their importance in industrialized countries also cannot be overstated. A study in [1] asserts that forests meet a variety of fundamental human requirements, including those for food, energy, fodder, housing, fiber, and medicine. They also aid in soil and water preservation, wildlife refuge, water purification, oxygen production, carbon sequestration, and climate regulation. In order to balance material prosperity, social welfare, and environmental health and vitality, it is necessary to acknowledge the significance of each of these functions and values in the forest ecosystem dynamics. Despite these benefits, there is a pressing need to categorize forests to uncover information about land cover, forest change detection, and which forest management activities to employ. Studies have shown that forests do not remain in the same state forever. They change their status due to global climate change, and natural and artificial disturbances [2].

Classifications of forest vegetation offer a practical method for summing together the information about the patterns of forest vegetation. This information is required to successfully plan for land use, map landscapes, and preserve natural habitats. Earth's ecological environment is hugely affected by so many factors and forest vegetation is key among them. Some of the ecological functions provided by forests include the protection of biodiversity, water conversation, and climate regulation [3] [4]. Because forest care is vital for the future, the United Nations proposed seventeen sustainable goals, where the 15th goal speaks about forest care, and this has necessitated studies in forests [5]. Forest departments survey land cover to obtain information such as the type of vegetation cover, quality and quantity of vegetation, and the dynamic changes that occur in forest land cover [6]. Remote sensing technology has provided high spatio-temporal resolution images with many spectral bands making conducting research in forestry easy. In that regard, artificial intelligence technologies, in particular, deep learning models have been applied in forestry to classify trees according to species [7], and to assess forest damage [8]. The Ministry of Forestry relies on remote sensing technology since it is the only efficient and cost-effective way to track forest vegetation's status and dynamic change across huge areas. The development and widespread usage of high-resolution remote sensing images like Quick Bird's submeter spatial resolution have provided a convenient way of studying forest vegetation in recent years. The field of remote sensing research is constantly adapting to keep up

with the demands of newly developed algorithms and increased computing power in parallel. Both the theory and the practice of remote sensing have significantly changed as a result of recent technological advancements, such as the creation of new sensors and improvements in data accessibility [9]. Applications related to land use and land cover (LULC) mapping are benefiting significantly from the increased availability of remote sensing images hence new opportunities to come up with innovative remote sensing classification techniques for various land management facets to address local, regional, and global challenges [10] [11] [12].

Forest experts play an important role in coming up with novel methods for forest image classification. They are actively involved in devising image analysis methods that are tailored to address the unique characteristics of forest ecosystems. Their domain knowledge helps identify key features and patterns within images that are indicative of forest attributes such as tree species, vegetation density, and land cover types [13].

Forestry experts participate in interpreting the functionality and relevance of the developed methods within the context of forestry management, conservation, and research. They evaluate the effectiveness of the methods in capturing meaningful information about forest conditions, identifying potential limitations, and providing feedback for further refinement [14].

Some novel approaches to forest image classification involve integrating data from different sources such as satellite imagery, LiDAR (Light Detection and Ranging) data, and hyperspectral imaging. Each of these data sources provides unique information about the forest environment. Satellite imagery offers broad spatial coverage but limited spectral resolution. LiDAR data provides precise elevation information and structural details of the forest canopy. Hyperspectral imaging captures fine-grained spectral information, allowing for more accurate discrimination of different forest types and conditions. Convolutional neural networks (CNNs) can be used to extract features from satellite imagery, while point cloud processing techniques can be applied to LiDAR data. Hyperspectral data can be processed using spectral analysis techniques to extract spectral signatures associated with different vegetation types and forest conditions. Once features are extracted from each modality of data, they are integrated into a unified feature representation that captures the complementary information provided by each data source. The integrated feature representation is then used as input to a machine learning classifier, such as a support vector machine (SVM), random forest, or deep neural network, to perform forest image classification. The classifier learns to discriminate between different forest types and conditions based on the combined information from multiple data sources.

Another novel approach to forest image classification incorporating deep learning and ontologies involves the integration of semantic knowledge representation with powerful feature learning capabilities [15]. This approach involves Utilizing deep learning techniques, such as convolutional

neural networks (CNNs), to extract high-level semantic features from forest images. Train CNN models on large-scale annotated datasets to learn discriminative features that capture spatial patterns, texture information, and spectral characteristics indicative of different forest classes. This is then followed by Integrating the deep learning-based features with semantic information from the forest ontology. This fusion process aligns the learned features with the semantic concepts defined in the ontology, enhancing the interpretability and contextual relevance of the extracted features. The final stage is to employ ontologically-driven classification algorithms that leverage the combined feature representations to classify forest images into meaningful semantic categories defined within the ontology. This classification process benefits from the semantic richness of the ontology, enabling more informed decision-making and improved generalization to unseen data.

1.2 Problem Statement

Data-driven methods, including supervised classifiers (such as Random Forest) and deep learning classifiers, are gaining much importance in processing big earth observation data due to their accuracy in creating observable images. Though deep learning models produce satisfactory results, researchers find it difficult to understand how they make predictions because they are regarded as black-box nature owing to their complicated network structures [16]. However, when inductive inference from data learning is considered, data-driven methods are less efficient in working with symbolic information. In data-driven techniques, the specialized knowledge that environmental scientists use to evaluate images obtained through remote sensing is typically disregarded. This limitation presents a significant obstacle for end users of Earth Observation applications who are accustomed to working with symbolic information, such as ecologists, agronomists, and other related professionals. Inference rule-based systems are used only in applications like the land cover classification systems developed by the FAO. It would appear that knowledge-driven approaches are the best way to go about conducting research in the field of remote sensing science. Because these rule-based approaches are founded on low-level feature image analyses, it is challenging for users to directly interpret and categorize images on their own. The focus should change from low-level to high-level image analysis which is flexible in dealing with symbolic information. This will make it easier for users to analyze and categorize images. Previous studies reveal that users are much more comfortable in seeking information using high-level semantics [17]. The main research question that guides this study is whether the integration of ontologies and deep learning approaches has a positive impact on the accuracy of forest image classification. In other words, the study aims to determine if the combined use of ontologies and deep learning can enhance the ability to accurately classify forest images into their different categories. The purpose of this study is to integrate deep learning

techniques with knowledge representations such as ontologies that harmonize high-level human knowledge to reduce the semantic gap between low-level image analysis and high-level human knowledge. There is currently no established, all-encompassing method for classifying images of forests in a way that is driven by semantics. There are still open questions about the best ontological framework to utilize for classifying forest images. What kind of ontological approach is appropriate for forest vegetation classification? How to present the ontology for forest vegetation classification? What is the efficient approach to harmonizing ontologies and deep learning techniques for forest vegetation classification? In this study, we provide a methodology framework for forest vegetation classification that combines ontologies with deep learning methods. The study also investigates the strengths and shortcomings of various techniques that have been employed to classify forest images. Thereafter, a hybrid approach that combines the strength obtained from the related studies is used to build an ontological approach for forest image classification.

Overall, the effectiveness of the models is assessed using metrics such as RMSE (measures the average magnitude of the errors between predicted values and observed values), MAE (measures the average absolute difference between the predicted values and the actual values in a dataset), confusion matrix (which is a table used to evaluate the performance of a classification model by presenting the model's predictions against the actual outcomes in a tabular format, ROC_AUC curves (is a graphical plot that illustrates the diagnostic ability of a classifier system as its discrimination threshold is varied), Accuracy (which is the proportion of correctly classified instances out of the total instances), Jaccard score index (A measure of similarity between two sets), precision (The proportion of true positive predictions out of all positive predictions made by the model), recall (The proportion of true positive predictions out of all actual positive instances) and F1-Score (The harmonic mean of precision and recall).

1.3 Aim and Objectives of the study

The study aims to develop a new ontological framework for forest image classification. The secondary objectives are to:

- Examine the current approaches used in forest image classification.
- Evaluate contemporary methodologies suitable for forest image classification.
- Create a structured framework tailored for segmenting forest imagery.
- Develop and implement a hybrid model combining deep learning and machine learning for classifying forest images.

- Develop a comprehensive framework integrating deep learning techniques with ontology for forest image analysis.

1.4 Thesis layout

The structure of the thesis is organized as follows: Chapter 2 reviews recent and appropriate methods for forest image classification. The chapter also presents a thorough study of the state-of-the-art model which has a high efficacy in classifying satellite forest images. Chapter 3 describes 2 proposed methods that hybridize convolutional neural networks and traditional machine learning classifiers for the purpose of segmenting satellite forest images. Chapter 4 describes 2 techniques employed to classify forest images. One approach is an ensemble of traditional machine learning and deep learning and the other technique is the state-of-the-art model that harmonizes deep learning and ontologies. Finally, Chapter 5 concludes the thesis with a summary of the presented methods as well as recommendations for future studies.

1.5 Methodology

1.5.1 Research Philosophy

This study uses the positivist research philosophy. The positivist approach to research is characterized by an emphasis on the utilization of quantitative techniques to put hypotheses to the test and determine the nature of the relationships that exist between different variables [18]. This study follows that reality is objective and that it is possible to observe and assess it using empirical evidence. This study uses an experimental design approach to determine the efficacy of a deep learning approach that uses ontology for forest image classification.

1.5.2 Research Method

The research method of this study consists of:

- examining the benefits, drawbacks, and complementing components of the existing forest image classification techniques.
- designing and defining a new methodology that (i) takes advantage of the strengths of current approaches, (ii) leverages cutting-edge methods for segmenting forest images, and (iii) applies ontology and deep learning methods to classify forest images.

1.5.3 Research Validation

The quantitative findings of this investigation are verified by means of both statistical analysis and replication. The validity of the experiment has been measured using a variety of metrics, including root-mean-square error, mean absolute error, Jaccard score index, accuracy, precision, recall, F1-score, and confusion matrix. Using the approach of replication, the same experiment was repeated several times over the same data and the results were consistent.

1.6 Thesis contribution

The contribution of our study toward forest image segmentation and classification is as follows:

- We carried out a critical analysis and survey of existing forest image segmentation and classification methods. The survey unearthed techniques used to detect objects of interest in an image that will be analyzed for the classification of forests. We also highlighted state-of-the-art models as well as performance and evaluation metrics for assessing forest image classification techniques.
- A review of modern ontology-based remote sensing applications for forest image classification gave an insight into the power of ontologies to explicitly represent knowledge, thereby, improving the classification process
- We designed a hybridized approach of deep learning models and machine learning techniques for forest image segmentation. A hybrid of ResNet50 and VGG16 was adopted to generate a set of features for the machine learning classifiers to perform the segmentation task.
- We developed a robust forest image classification model by combining the deep learning technique ResNet50 and the machine learning model XGBoost. Deep learning was selected because it excels at feature extraction, while XGBoost was chosen because it excels at image classification.
- We demonstrated that the integration of semantic ontologies in deep learning models helps to eliminate error propagation through weak attribute learning, and this concept increased image classification accuracy.
- We demonstrated that the concept of aggregating outputs from hypernym-hyponym classifiers significantly forest image class or category distinction.

1.6.1 Chapters summary

Six separate sections make up this research. Section 1 provides a basic introduction with regard to ontology and forest image classification and the general outline of the thesis. Section 2 provides an analysis of the currently available techniques for classifying forest images as well as a discussion of appropriate techniques for classifying forest images, focusing on GEOBIA and ontology in particular. Section 3 focused on segmentation, a crucial step in the image classification process. Section 4 explains models that were used to classify forest images into their respective categories.Section 5 discussed overall results with respect to findings from the literature survey, segmentation, and classification. Lastly, Section 6 provides the study's conclusion and future recommendations. Figure 1 shows the relationship between the several papers as previously described. Additionally, Table 1 establishes the connections between the papers that were used to address the study's objectives (See Section 1.3).

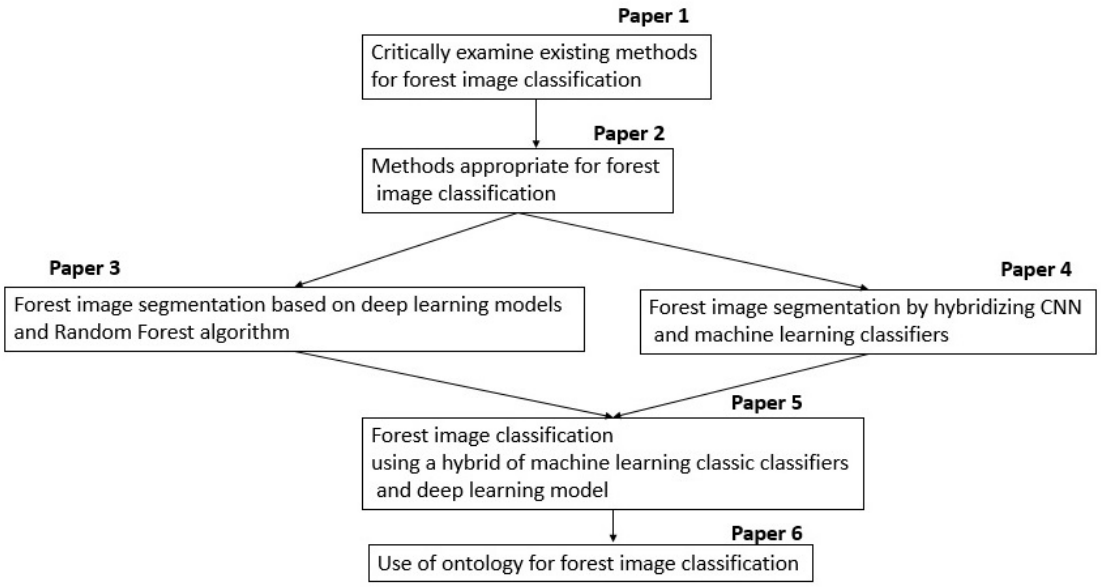


Fig. 1: A Graphical representation of the relationships between the papers that make up this thesis.

SECTION II

Machine Learning Methods for Forest Image Analysis and Classification: A Survey of the State of the Art IEEE Access. 2022 Apr 25;10:45290-316.

In this research study, a survey of the various knowledge-driven and machine-learning approaches that are currently applicable to forest image processing was undertaken. The study discussed the limitations of the image processing technologies that are now in use and recommended the use of

Table 1: Objectives and papers used to address the objectives

Objectives (1-5)	Papers used to address the objectives
1 (main Objective)	Paper 6
2	Paper 1
3	Paper 2
4	Paper 3 and 4
5	Paper 5

ontologies for the classification of forest images as a potential course of action for the future.

A Critical Survey of GEOBIA Methods for forest Image Detection and Classification. This article was published in GEOCARTO INTERNATIONAL journal (Taylor and Francis)

This article promotes the use of ontologies for knowledge representation in remote sensing. In order to do this, an overview of GEOBIA studies for image analysis and classification is provided, and the shortcomings of current approaches in meeting the expectations of remote sensing experts are made clear. The potential for improving GEOBIA models as well as new GEOBIA development methods have both been investigated. To demonstrate the benefits of ontologies in detecting and classifying satellite images, recent research that employed the concept of ontology in the classification of forest images is examined and recommendations for the remote sensing science community are given.

SECTION III

Hybrid Learning Model for Satellite Forest Image Segmentation. This article was published in In International Conference on Artificial Intelligence and Soft Computing (pp. 37-47). Cham: Springer Nature Switzerland.

In this research, we used a supervised method for segmenting satellite images of forests. The suggested model obtained a feature vector via ResNet50 transfer learning and then fed those features into a Random Forest for classification. The Land Cover Classification Truck in Deep- Globe Challenge provided the satellite images utilized in training and testing. The effectiveness of the model was assessed using metrics such as precision, recall, F1-Score, accuracy, Root Mean Square Error (RMSE), and Mean Average Error (MAE).

Hybridizing Deep Neural Networks and Machine learning Models for Aerial Satellite Forest Image Segmentation. Under Review

Forests are a vital natural resource because they substantially mitigate climate change and contribute to the economic growth of many nations. In this regard, it is essential to constantly monitor forest cover. This paper proposes a hybridization of deep neural networks (VGG16 and ResNet50) and machine learning classifiers for segmenting a forest image into forest and non-forest regions. The sole purpose of deep neural networks was to produce a set of features for machine learning classifiers to use in the segmentation process. In this study, RF, LSVM, LDA, GNB, and kNN were used as machine learning classifiers. Accuracy, the Jaccard score index, root-mean-square error, precision, and recall were used to evaluate the model's performance.

SECTION VI

Forest Image Classification based on Deep Learning and XGBoost Algorithm. This article was published in the International Conference on Computational Science (pp. 217-229). Cham: Springer Nature Switzerland.

This research aims to create a hybrid model that combines deep learning and machine learning to enhance the precision of forest image classification. In order to improve the prediction capability of classifying satellite forest images, this study proposes an ensemble approach using the Deep Learning technology (ResNet50 in particular) and machine learning model (particularly XGBoost). ResNet50's primary function is to produce a feature set for use in the subsequent classification work done by the XGBoost algorithm. Classifiers like random forest (RF) and light gradient boost machine (LGBM) were used to evaluate the XGBoost algorithm alongside a fully connected ResNet50 model.

Ontology with Deep Learning for Forest Image Classification. Applied Sciences 13, no. 8 (2023): 5060.

This study constructed the ontology using concepts of image categories and the taxonomic link between them. Graphical semantic information describing the training images was provided by the

ontology. Each super-image category's features were used to train a hypernym classifier recursively. Finally, both the hypernym and hyponym classifiers were used to sort the test images into their designated categories. It's worth noting that compared to the control approaches, the proposed model of harmonizing deep learning models and ontologies achieved better results. In the field of forestry, the ontological bagging method can be used to categorize trees by species and other forms of vegetation.

References

[1] E. K. Nunoo, "Eia performance standards and thresholds for sustainable forest management in ghana," *Standards and thresholds for impact assessment*, pp. 229–240, 2008.

[2] A. Sommerfeld, C. Senf, B. Buma, A. W. D'Amato, T. Després, I. Díaz-Hormazábal, S. Fraver, L. E. Frelich, Á. G. Gutiérrez, S. J. Hart *et al.*, "Patterns and drivers of recent disturbances across the temperate forest biome," *Nature communications*, vol. 9, no. 1, p. 4355, 2018.

[3] Z. Biao, L. Wenhua, X. Gaodi, and X. Yu, "Water conservation of forest ecosystem in beijing and its value," *Ecological economics*, vol. 69, no. 7, pp. 1416–1426, 2010.

[4] E. Führer, "Forest functions, ecosystem stability and management," *Forest Ecology and management*, vol. 132, no. 1, pp. 29–38, 2000.

[5] X. Cheng, A. Doosthosseini, and J. Kunkel, "Improve the deep learning models in forestry based on explanations and expertise," *Frontiers in Plant Science*, vol. 13, 2022.

[6] G. Jiang and Q. Zheng, "Remote sensing recognition and classification of forest vegetation based on image feature depth learning," *Mobile Information Systems*, vol. 2022, 2022.

[7] F. H. Wagner, A. Sanchez, Y. Tarabalka, R. G. Lotte, M. P. Ferreira, M. P. Aidar, E. Gloor, O. L. Phillips, and L. E. Aragao, "Using the u-net convolutional network to map forest types and disturbance in the atlantic rainforest with very high resolution images," *Remote Sensing in Ecology and Conservation*, vol. 5, no. 4, pp. 360–375, 2019.

[8] H. Tao, C. Li, D. Zhao, S. Deng, H. Hu, X. Xu, and W. Jing, "Deep learning-based dead pine tree detection from unmanned aerial vehicle images," *International Journal of Remote Sensing*, vol. 41, no. 21, pp. 8238–8255, 2020.

[9] D. Arvor, M. Belgiu, Z. Falomir, I. Mougenot, and L. Durieux, "Ontologies to interpret remote sensing images: why do we need them?" *GIScience & remote sensing*, vol. 56, no. 6, pp. 911–939, 2019.

[10] S. K. Karan and S. R. Samadder, "Improving accuracy of long-term land-use change in coal mining areas using wavelets and support vector machines," *International Journal of Remote Sensing*, vol. 39, no. 1, pp. 84–100, 2018.

[11] J. J. Gapper, H. El-Askary, E. Linstead, and T. Piechota, "Coral reef change detection in remote pacific islands using support vector machine classifiers," *Remote Sensing*, vol. 11, no. 13, p. 1525, 2019.

[12] T. Bai, K. Sun, S. Deng, D. Li, W. Li, and Y. Chen, "Multi-scale hierarchical sampling change detection using random forest for high-resolution satellite imagery," *International Journal of Remote Sensing*, vol. 39, no. 21, pp. 7523–7546, 2018.

[13] J. W. Atkins, P. Bhatt, L. Carrasco, E. Francis, J. E. Garabedian, C. R. Hakkenberg, B. S. Hardiman, J. Jung, A. Koirala, E. A. LaRue *et al.*, "Integrating forest structural diversity measurement into ecological research," *Ecosphere*, vol. 14, no. 9, p. e4633, 2023.

[14] L. G. Bont, M. Fraefel, F. Frutig, S. Holm, C. Ginzler, and C. Fischer, "Improving forest management by implementing best suitable timber harvesting methods," *Journal of Environmental Management*, vol. 302, p. 114099, 2022.

[15] C. Kwenda, M. Gwetu, and J. V. Fonou-Dombeu, "Ontology with deep learning for forest image classification," *Applied Sciences*, vol. 13, no. 8, p. 5060, 2023.

[16] D. Castelvecchi, "Can we open the black box of ai?" *Nature News*, vol. 538, no. 7623, p. 20, 2016.

[17] L. Yuee, "Ontology-based image annotation," Ph.D. dissertation, Queensland University of Technology, 2010.

[18] H. K. Mohajan *et al.*, "Quantitative research: A successful investigation in natural and social sciences," *Journal of Economic Development, Environment and People*, vol. 9, no. 4, pp. 50–79, 2020.

2 Previous Works: Survey Papers

2.1 Machine Learning Methods for Forest Image Analysis and Classification: A Survey of the State of the Art

2.1.1 Introduction

This section introduces a research paper that examines previous but recent technologies in satellite forest image classification. The paper analyzed the challenges of object detection methods, provides a solution framework for future direction as well as presents a state-of-the-art learning model for classifying remote sensing images.

This paper has been published by IEEE open-access journal.

Received March 9, 2022, accepted April 15, 2022, date of publication April 25, 2022, date of current version May 3, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3170049

Machine Learning Methods for Forest Image Analysis and Classification: A Survey of the State of the Art

CLOPAS KWENDA, MANDLENKOSI GWETU, AND JEAN VINCENT FONOU DOMBEU
School of Computer Science, Statistics and Mathematics, University of KwaZulu-Natal, Pietermaritzburg 3209, South Africa
Corresponding author: Clopas Kwenda (221072651@stu.ukzn.ac.za)

ABSTRACT The advent of modern remote sensors alongside the development of advanced parallel computing has significantly transformed both the theoretical and real implementation aspects of remote sensing. Several algorithms for detecting objects of interest in remote sensing images and subsequent classification have been devised, and these include template matching based methods, machine learning and knowledge-based methods. Knowledge-driven approaches have received much attention from the remote sensing fraternity. They do, however, face challenges in terms of sensory gap, duality of expression, vagueness and ambiguity, geographic concepts expressed in multiple modes, and semantic gap. This paper aims to review and provide an up-to-date survey on machine learning and knowledge driven approaches towards remote sensing forest image analysis. It is envisaged that this work will assist researchers in coming up with efficient models that accurately detect and classify forest images. There is a mismatch between what domain experts expect from remote sensing data and what remote sensing science produces. Such a mismatch or disparity can be reduced or alleviated by adopting an ontology paradigm methodology. Ontologies should be used to support the future of remote sensing in forest object classification. The paper is presented in five parts: (1) a review of methods used for forest image detection and classification; (2) challenges faced by object detection methods; (3) analysis of segmentation techniques employed; (4) feature extraction and classification; and (5) performance of the state-of-the-art methods employed in forest image detection and classification.

INDEX TERMS Feature extraction, ontology, segmentation, remote sensing.

I. INTRODUCTION

Remote sensing science is rapidly growing. The evolution of high spatial resolution remote sensors in conjunction with advanced computing has significantly transformed the specification and practice of remote sensing [1]. Remote sensing images are characterized by high spatial resolution and provide more explicit information on the earth's surface as compared to middle and coarser resolution images [2]. Machine learning methods for analyzing and classifying forest images are continuously evolving to provide more advanced automatic land cover pattern recognition on aerial images.This paper surveyed existing methods for forest ecosystem image classification. In particular, machine learning classifiers and deep learning techniques for forest image classification are reviewed.

There are several algorithms that are geared towards detecting objects of interest in remote sensing images, for the further regional analysis and classification. These algorithms are categorized into three groups, namely template matching based, knowledge based, and machine learning based methods [3]. The taxonomy of image classification methods is depicted in Figure 1.

(a) Template matching based detection methods

The template matching method determines whether a picture or an image contains a previously defined object or whether a predefined sub image (template) has an exact match in an image. Although this method provided one of the first approaches for object analysis [3], its dependence on handcrafted matching criteria limited its applicability to complex object recognition. Once a suitable template is deter-

The associate editor coordinating the review of this manuscript and approving it for publication was Stefania Bonafoni.

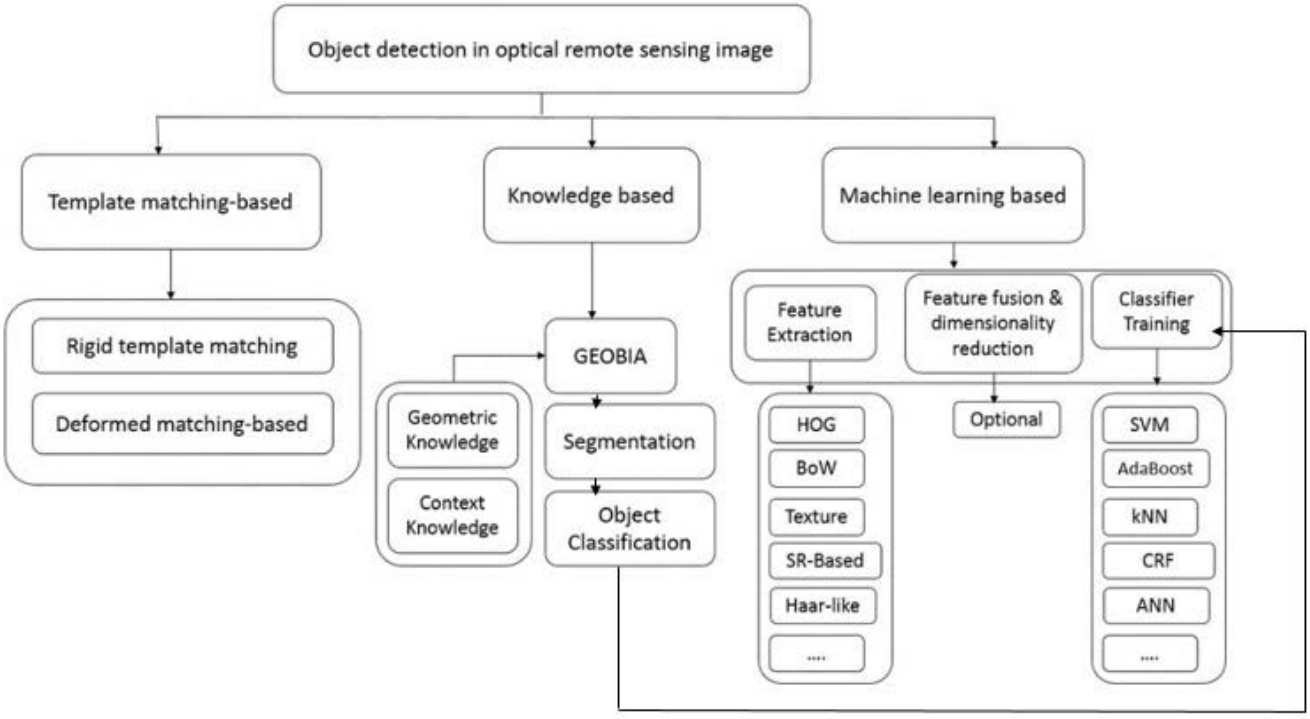


FIGURE 1. Methods for object detection in optical sensing images [3].

mined, a measure of matching between the template and every possible location in the image is calculated, and a classification decision is made based on the measure of certainty. The most popular metric based measures are the Euclidean distance, squared difference, and cross correlation, defined in Equations (1) to (3):

Euclidean distance

$$E\,(m,n) = \sqrt{\sum_i \sum_j [g\,(i,j) - t\,(i-m, j-n)]^2 24} \qquad (1)$$

Squared Difference Measure

$$E^2\,(m,n) = \sum_i \sum_j [g(i,j)^2 - 2g(i,j)t(i-m,j-n)) + t(i-m,j-n)^2] \qquad (2)$$

Cross Correlation Measure

$$R\,(m,n) = \sum_i \sum_j g\,(i,j) + t\,(i,j)\,t\,(i-m,j-n) \qquad (3)$$

There are two types of templates, namely, global and local templates. When a template is used to reference the whole (global) object in an image, it is referred to as a global template. However, when object features (local features of an object) in an image are referenced with multiple or several templates, these templates are referred to as local templates [4]. Figure 2 shows stages to be followed to determine the best templates for object detection. The challenge

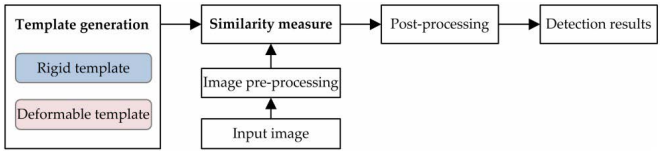


FIGURE 2. Template matching based criteria [3].

with this approach is that the method does not cater for the scale and orientation of the template [5]. It fails due to occlusions and distortions on the boundary [6]. The method is very sensitive to shape and viewpoint change. The solution suggested was to have a unique representation of a template orientation and scale that varies, but the solution becomes computationally expensive.

(b) Machine Learning based approaches

An input image is subjected to the initial first phase where regions or objects are extracted. Then, for each object, features of interest are computed using Convolutional Neural Networks (CNN). Optimal features are obtained after a subsequent series of feature fusion and dimension reduction processes. Finally, classifiers such as Support Vector Machines (SVM), k-Nearest Neighbor (kNN), Sparse Representation based Classification (SRC), AdaBoost, Conditional Random Fields (CRF), and others are used to classify each region/object. Figure 3 shows the main important phases of machine learning object detection, i.e., feature extraction, feature fusion, and dimensionality reduction, as well as

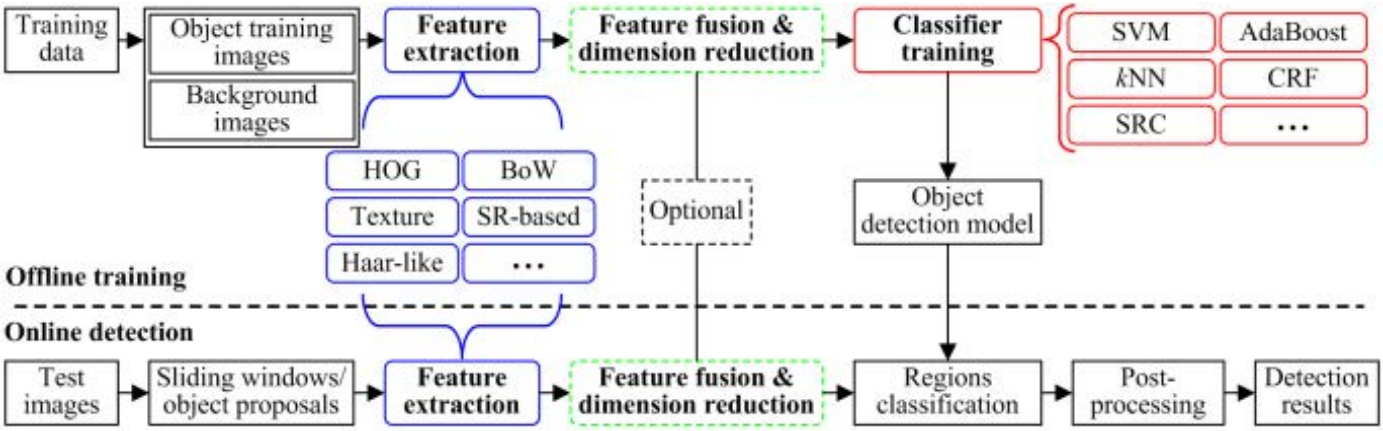


FIGURE 3. Machine learning methods [3].

the classification phase. Machine learning-based approaches coupled with innovative algorithms and higher performance computing seem to have gained popularity in remote sensing science because they produce better results considering the accuracy of the created maps [1]. As a result, they are used in big land cover applications that rely on pixel based statistical analysis of massive image data sets [7]–[9]. Pixel based approaches pose challenges in the analysis of high spatial resolution images [10] because they take into consideration the aspects of spectral information as a backdrop for analyzing and classifying high spatial resolution images, neglecting spatial and temporal information, which are of paramount importance. These methods are less efficient in dealing with symbolic knowledge, that is, when concepts are characterized by symbols, for instance, vegetation is made of grass [11]. They do not offer the function of creating a super class whilst classes have been defined. Suppose one has defined the following classes of interest; ''trees'', ''grass'', ''road'', and ''building''. It will then be impossible for the user to define a vegetation class unless it has been beforehand defined as a super class. The methods do not offer the facility to add spatial rules [12], for instance, ''grass'' cannot be found inside a building, but can be found in a field. Because of this reasoning limitation, data-driven approaches are unsuitable for use in applications areas such as ecology, that deal with earth observation.

(c) Knowledge based detection methods

These methods have been applied to land slide, crops, urban land change and forests [13]–[16]. Figure 4 shows the processes whereby an input image goes through a hypothesis generation phase, the hypothesis is validated and tested using the established knowledge and rules. Post processing of validation results will be subjected to machine learning for final object detection. Knowledge and rules from geometric information and context information will be used to test the validity of the hypothesis generated from an input image, if the hypothesis is valid it will be subjected to machine learning for object detection [17]. Generally, there are two

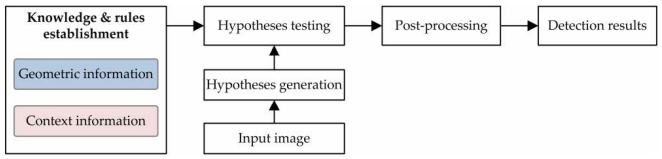


FIGURE 4. Knowledge based object detection systems [3].

types of knowledge that have been used on target objects, and these are geometric knowledge and context knowledge.

(a) Geometric Knowledge

This type of knowledge is the most important in a knowledge-based approach and is widely used for object detection. It encompasses generic shape models or parametric specifics. For instance, it is proposed in [18] that buildings are square or composed of rectangular segments and are utilized as conventional models of shapes to distinguish buildings.

(b) Context knowledge

Context knowledge is very important for key objects, and it is expressed by rules derived from relationships between objects of interest and their respective backgrounds [14], [19], [20]. For instance, shadow evidence has been used for building detection [21], the correlation between artificial structures such as buildings and their respective shadows has been used to project locations and shapes of buildings [22].

Recently, knowledge-driven approaches seem to be the direction taken by the remote sensing science community [3] since they incorporate domain expert knowledge. Geographic Object Based Image Analysis (GEOBIA), which classifies image objects based on apriori domain expert knowledge, is proving to be a key trend in remote sensing image analysis. GEOBIA is a classification technique that divides a remote sensing image into objects of interest and evaluates the objects based on their spectral, temporal and spatial characteristics. The generation of objects of interest is done using different segmentation approaches such as random walker, canny, histogram-based segmentation, etc. An algorithm is deemed effective in segmentation if and only if a segmented

image object completely matches the corresponding Actual Image Region (AIR) of a scene object. [2], proposed a blend of area coincidence methods and boundary coincidence methods for assessing segmentation quality. The area coincidence methods select an image that has the dominant or largest area of intersection with the AIR. The boundary coincidence methods calculate the distance between a point of interest in a segmented image and that of its corresponding point on the AIR. The segmentation quality is high when the measured distance is much closer to zero. Segmentation evaluation methods can either be Unsupervised or Supervised. Supervised techniques evaluate a segmented image based on a ground truth image also referred to as the reference image. The evaluation of unsupervised methods is solely dependant on the segmented image as it has to assess the extent to which the image matches the desirable features of a good segmented image. [23] proposed four metrics (Equations 4-7) for assessing supervised segmentation quality namely F-measure, SUM which should be less than 2, ED that indicates distance to point (0,0) in the space and ED' that indicates distance to point (1,1) in the space.

$$f = \frac{1}{\alpha \frac{1}{precision} + 1 - \alpha \frac{1}{recall}} \tag{4}$$

$$sum = precision + recall \tag{5}$$

$$ed = \sqrt{precision^2 + recall^2} \tag{6}$$

$$ed' = \sqrt{(1 - precision)^2 (1 - recall)^2} \tag{7}$$

Two other metrics that take into account the over and under segmentation errors, GOSE and GUSE, respectively, are proposed [24]. Rand Error(RE) is another widely used metric for evaluating supervised approaches. RE is a measure of is defined in Equation 8 [25]. Let R_1 and R_2 be segmentation regions of image S with t pixels and the following holds:

- n correspond to the number of pixels in image S that appear in both R_1 and R_2
- m correspond to the number of pixels in image S that are neither in R_1 and R_2

$$RE = \frac{n + m}{\binom{n}{2}} \tag{8}$$

A criterion for unsupervised technique that balances homogeneity and inter-segment heterogeneity is proposed by Wang *et al.* [26] as in Equation 9.

$$Z = T + \lambda D \tag{9}$$

where, T and D represents intra-segment homogeneity and inter-segment heterogeneity, respectively. Another metric for unsupervised techniques proposed by Gao *et al.* [27] is the Global Score (GS). GS incorporates weighted variance (WV) and Moron's I and is defined in Equation 10.

$$GS = V_{norm} + I_{norm} \tag{10}$$

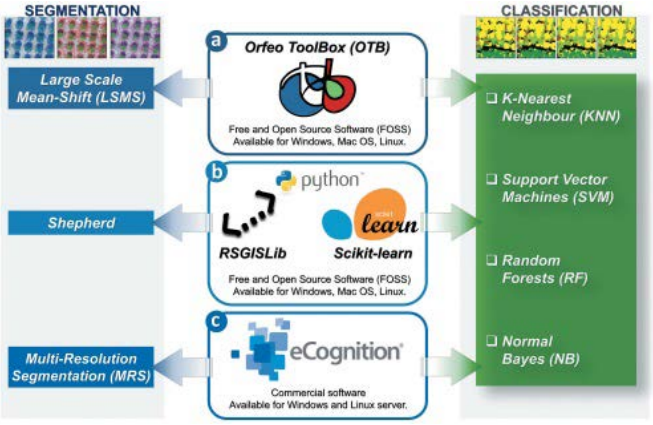


FIGURE 5. GEOBIA WORKFLOW [28].

The final step of GEOBIA is image classification. The common image classifiers for GEOBIA are Random Forest (RF), Simple Vector Machines (SVM), k-Nearest Neighbor (kNN) and Naive Bayes (NB) [28].

Figure 5 shows GEOBIA workflow [28] that implemented three different algorithms, namely, Large Scale Mean Shift (LSMS) in OTB, the Shephered segmentation algorithm in RSGISLib and the Multi-resolution segmentation (MRS) algorithm in eCognition. However, GEOBIA solutions do not give answers to every segmentation problem. Even though GEOBIA is more efficient than pixel-based approaches, segmenting a multi-spectral image made up of thousands of mega pixels remains a challenging task [29]. Another drawback of GEOBIA is that it approximates, to some extent, computer-aided photo interpretation, which has been criticized as being highly subjective [5]. However, in the last decade, knowledge-driven techniques, like GEOBIA, have gained traction as a means of bridging the gap between implicit data representation and end-user needs. Knowledge-driven approaches consist of translating symbolic knowledge into a format understandable by humans into numerical knowledge.

Vegetation indices obtained from satellite images provide valuable information which is essential for the mapping of vegetation. The Normalised Difference Vegetation Index (NDVI) has proven to be a valuable tool, particularly in tropical dry forests, where it serves as a foundation for estimating overall green biomass, tree density, and species diversity [30]–[32]. NDVI is an indicator that determines the greenish component from the analyzed satellite images. NDVI provides a balance between the energy received and the energy emitted by objects on the earth's surface [32]. In the context of plant communities, it is an indicator that determines how greenish an area is, and that is influenced by the quantity of vegetation in that particular area and its state of health. The NDVI values range from -1 to $+1$. The values that are less than 0.1 correspond to water bodies and bare grounds, while higher values indicate the presence of agricultural activities, temperate forests, and rain

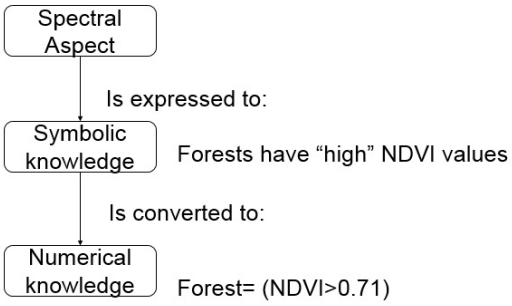


FIGURE 6. Symbolic to numerical knowledge conversion.

forests [32]. The NDVI values can be used to group the vegetation ecosystems into 4 major categories as follows [33]: forests made up of semi-deciduous and evergreen have NDVI $\geq$ 0.7, woodlands are defined by the range (0.6 $\leq$ NDVI $<$ 0.71), a mixed class that is composed of a) shrub land, b) woodland/shrub land/exposed lands, and c) cactus forest have the range (0.49 $\leq$ NDVI $<$ 0.61) and the dwarf woodland and shrub land assume the range of (NDVI $<$ 0.49) [33]. For instance, a forest concept made up of semi-deciduous and evergreen is described by high NDVI values and when translated into numerical knowledge, it is implemented by the classification rule set: Forest = (NDVI $\geq$ 0.71). Figure 6 shows the symbolic to numerical knowledge conversion.

II. OVERVIEW OF THE CHALLENGES FACED WITH KNOWLEDGE BASED APPROACHES

A. DIFFERENT MODES OF DEFINING GEOGRAPHIC CONCEPTS

A geographical concept can be defined from different perspectives; the definition might be based on physical, historical, functional, or conventional mode [34]. Various methods of defining the same geographic concept bring about elective perspectives on the definition of the same concepts; for example, an idea can be characterized by elective definitions that are not basically the same, despite the fact that they are normally correlated [35]. From a functional viewpoint, the role of the forest primarily acts as a repository for storing carbon. This is correlated by the Net Primary Productivity (NPP) values. Forests can also be defined based on physical attributes such as vegetation cover, phenology, vegetation, age, etc. A tremendous effort is still in place to standardize land cover classes in land cover classification systems (LCCS) [36]. The term ''forest'' is defined differently by different organizations and countries; for instance, in Brazil an area that is regarded as a forest, has an area that exceeds one hectare, is characterised by a 30% canopy, and is composed of trees with a minimum height of 5m [37]. A forest in China is defined as an area larger than 0.67 hectares in size, with at least 20% crown cover and trees standing at least 2 meters tall. The Food and Agriculture Organization (FAO) standardised the definition of forest to refer to a land area spanning over 0.5 hectares enveloped by trees at 5m and above, with a canopy cover of 10%. This definition excludes land under agricultural or urban land use [38].

B. DUALITY OF GEOGRAPHIC CONCEPTS

Two important major terms arise from the concept of duality, that is, scene and image. A scene is real and exists on the ground, whereas an image is an assortment of spatially orchestrated estimations drawn from the scene [1]. Components obtained from images are regarded as abstractions of real objects in the ground scene [39]. Forest concepts can be viewed either from a real world perspective (a forest concept is characterized by high NPP values) or from image properties (a forest concept is defined by high NDVI values). In the case of forests, the assertion that NDVI is correlated to NPP is not always valid because NPP lacks information on attributes such as vegetation height, vegetation cover, and so forth. This anomaly is also referred to as the sensory gap. With this notion, sensory gaps cause improbability in describing geographic objects [40].

C. VAGUENESS AND AMBIGUITY OF GEOGRAPHIC CONCEPTS

The process of connecting attribute (e.g. NDVI) range of values (for instance, high) to geographic concepts (e.g. forest concept) is not easy [1]. The reason behind this is that the associated value ''high'' is qualitative in nature, so the obtained classification rule becomes vague and ambiguous. Some pixels (image objects) inside the image in Figure 7 are not classified as forest, though in nature they are constituted as forest areas. When qualitative terms like ''high'' are employed to identify objects with sharp, crisp boundaries, threshold ambiguity occurs. Qualitative description of geographic objects raises partiality issues. For instance, the symbolic classification rule ''high NDVI'' partially describes the forest because:

(a) It is very difficult to fish out only forests in areas that have other crops with the same NDVI values as in Figure 9.

(b) It is also impossible to classify all the varying types of forests because, in some cases, there are some forests that have ''low NDVI'' values, such as the degraded forests in Figure 8.

Ambiguity arises in all cases where a natural language expression can have various meanings [41]. The usual one is lexical ambiguity, which emerges because of the homonym of regular language articulation, that is, an expression with more than one meaning, such that each meaning points only to one ontological concept unambiguously [41]. More than 800 different definitions of forest concepts are provided in [42]. Deep ambiguity, also referred to as open texture, exists where there is no clear boundary between concepts or terms or cases where the meaning of a concept changes over time, for

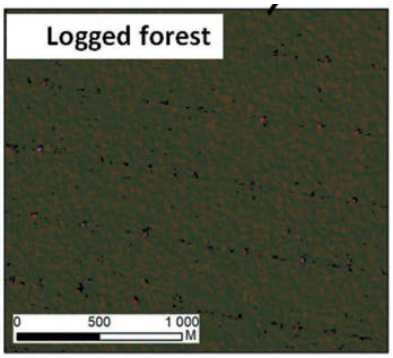


FIGURE 7. Concepts of vagueness.

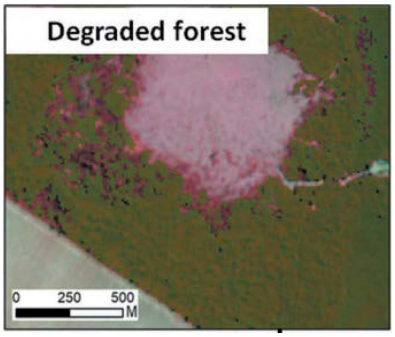


FIGURE 8. Forests having low NDVI values.

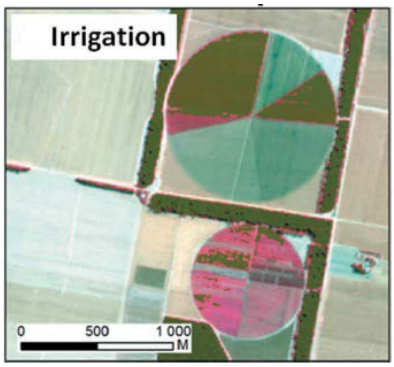


FIGURE 9. Other crops having high NDVI as forests.

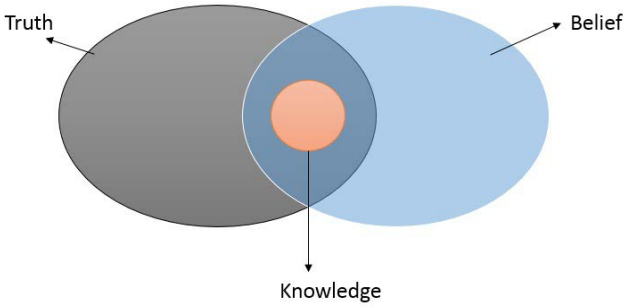


FIGURE 10. Definition of knowledge [45].

instance, when a new technology appears or the physical or social context of the term evolves.

D. SEMANTIC GAP

It arises from the vagueness and ambiguity of geographic concepts. It is defined as a mismatch between data extracted on the basis of visual information and the interpretation drawn from the same data in a given situation [40]. This is so because converting visual data (from human perception) to computational representation is a very difficult task. The translation first requires expressing perception of visual data into a symbolic knowledge representation format (e.g. forests have high NDVI values). Such a conversion is a very difficult task since some concepts have vague meanings when expressed in natural language [43]. For instance, color may be considered a significantly important biophysical property [44], but its perception varies amongst humans and it is difficult to express.

III. INTRODUCTION ON ONTOLOGIES

Sharing knowledge among people is feasible only if people speak a common language [42]. The traditional definition of "knowledge is a subset of true beliefs" [45]. It is the intersection between truth and beliefs, as represented in Figure 10. Ontologies enable formal (machine-understandable) representation of knowledge. In computer science, ontology is defined as an explicit, formal specification of a shared conceptualization [46]. An ontology is a systematic description of existence, and this term is drawn from philosophy. What "exists" for Artificial Intelligence (AI) systems is that which can be represented. The following properties, with corresponding definitions, should be observed: (1) conceptualization, means that an ontology is an abstract model of a real world phenomenon; (2) explicit, implies that all ontology concepts must be clearly defined; (3) formal, implies that an ontology is machine understandable; and (4) shared, means that there should be consensus amongst a community of people about the knowledge represented by the ontology.

A. FORMAL ONTOLOGIES

Remote sensing science experts are conversant with working on numerical knowledge that has been derived from an image viewpoint [47]. Numerical knowledge representation by nature suffers from the problem of partiality and implicit knowledge representation, hence it becomes difficult to share the knowledge with other scientists, such as ecologist, agronomist, etc., who are used to working with symbolic knowledge in describing a geographic concept, for instance, a forest concept is defined by "High NDVI" values. Formal ontologies provide a road map that caters for the representation of both symbolic and numerical knowledge. Formal ontologies can be utilized to unequivocally portray a perception or observation from different perspectives, for instance, the extensible observation ontology (OBOE) is utilized to portray the semantics of scientific observations. An observation of an entity encompasses the characteristics (e.g. biomass) of the entity based on a measurement standard

2. PREVIOUS WORKS: SURVEY PAPERS

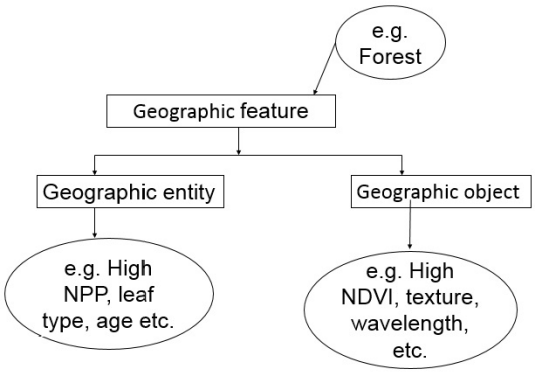


FIGURE 11. Dual representation of concepts.

(grams). Ontologies for remote sensing science applications based on description logic offer the following advantages;

- Symbolic language - it binds/associates concepts with relevant sensing data and also promotes binding of related concepts.
- Knowledge sharing - it advocates for common conceptualization and adoption of standard ontology language such as web ontology language.
- Reasoning - description logic in ontology allows the inferring of new knowledge from explicit descriptions.

B. ONTOLOGY KNOWLEDGE BASED AS A SOLUTION

This section outlines how the adoption of ontologies in knowledge base approaches helps in alleviating the problems addressed in Section 2.0.

1) DUALITY OF GEOGRAPHIC CONCEPTS

Ontologies incorporate the concept of perspectivalism. That is, they allow the separate description of a field point of view of a forest concept. For instance, a forest concept can be specified in terms of attributes such as ''high'' NPP, leaf type, and so on. The other angle of description is from the point of view of an image of geographic objects. That is, a forest can be defined in terms of attributes such as ''high'' NDVI, texture, and wavelength. In general, it allows for the separate description of geographic entities and geographic objects alongside their characteristics. Figure 11 shows the dual representation of a geographic feature, that is, it can be described either from the perspective of a geographic entity or from the perspective of a geographic object.

2) VAGUENESS AND AMBIGUITY OF GEOGRAPHIC CONCEPTS

Fuzzy logic is the most popular way of handling the vagueness of geographic concepts [48]. Processing of data is done using partial set membership rather than strict set membership. For example, a forest concept is not considered to be strictly ''green'', but rather is considered to belong partially to some degree to the set of things that are green. [49] defined

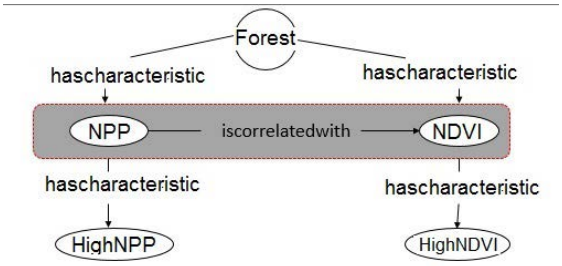


FIGURE 12. Solving sensory gap challenge.

two thresholds i.e. ambiguity reject threshold and the distance reject threshold. Ambiguity reject threshold is defined by the rule $\alpha_{amp} \in \mid 0.5, 1 \mid$ and define the degree of confidence required to recognise an object. Distance reject threshold is defined by the rule $\alpha_{dist} \in \mid 0.1 \mid$, this means an object x is unlikely to belong to both classes C_k and $\neg\ C_k$ and might belong to a concept not yet learnt. Vagueness can also be addressed by adopting probability ontologies. They use probability sets to define concepts of interest. Attributes in the set properties have probabilities attached to them, and the statistical measure of the probability value of the geographic concept [50] is used to determine whether a geographic concept is a member of a class. Ambiguity in ontologies can be reduced by limiting the information that describes a concept [41].

3) SENSORY GAP

The discrepancy between real objects and their depiction in images is known as the sensory gap. As referenced by [40] sensory gap can be reduced by explicitly defining the domain and world knowledge in the system. Knowledge about physical laws, laws governing the behavior of objects, and how people perceive them will all be incorporated into the system in the hope of enhancing recognisers and thereby assisting machines in bridging the sensory gap [51]. However, in ontologies, real world description of forest entities is correlated with matching image point description of forest objects, i.e., NDVI is correlated with NPP [2]. Figure 12 shows how a real world description of a forest concept can be mapped to an object description in an image. An object description of an image is easily formalised on a computer.

4) SEMANTIC GAP

The semantic gap is the discrepancy between the high level descriptions of images by humans and the low-level detection used by machines to detect images [52]. On the other hand, adding captions and annotations to images solves the problem [50]. The method is time-consuming and costly because it requires a lot of effort, machine algorithm tweaking, and close attention to vocabulary and content to ensure that photos are appropriately labeled [50]. In ontologies, however, an image feature (e.g. NDVI) and its associated value (''HIGHNDVI'') are used to define a pixel (image object) of a forest concept. The ''HIGHNDVI'' concept is formalized as a result

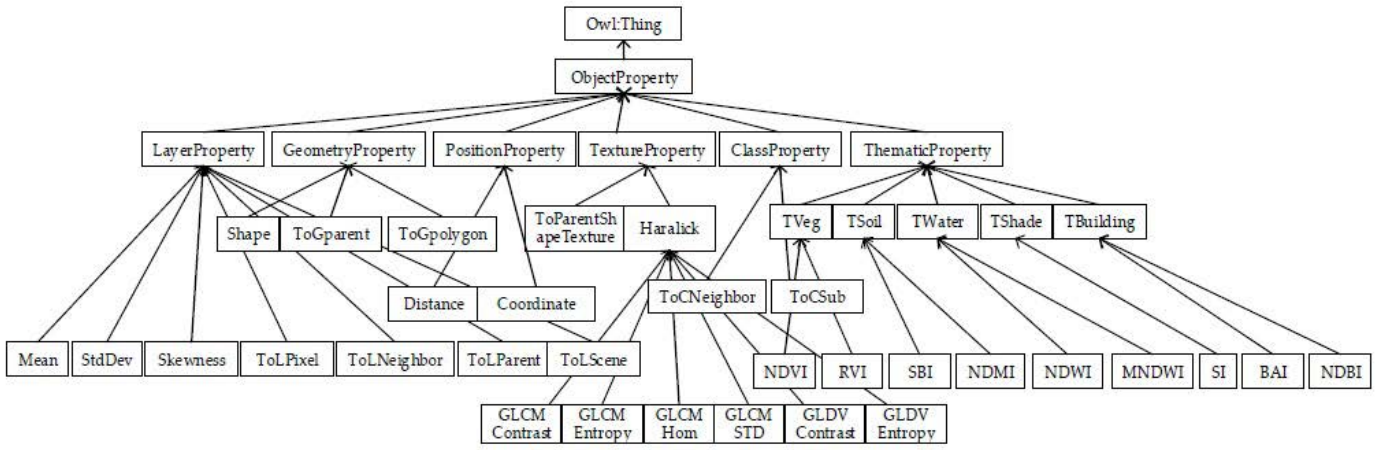


FIGURE 13. Object features hierarchy in ontology [55].

TABLE 1. Ontological framework for RSI concepts.

Item	Quantity	Image
Associated_to	Region	Image
from_band	Image	-
from_satellite	Sensor	Satellite
from_sensor	-	Sensor
has_spatial_resolution	-	spatial resolution
has_spectral_resolution	-	spectral resolution

of the established relationship between symbolic information (e.g. ''HIGHNDVI'') and numerical knowledge (e.g., NDVI>0.7), hence the semantic gap is reduced.

IV. ONTOLOGICAL FRAMEWORK FOR RSI (REMOTE SENSING IMAGE)

[53] proposed a novel framework for RSI. The framework is made up of important terms or concepts. These include satellite, sensor, image, spatial resolution, and spectral resolution. The elements are shown in the table 1. Slot is mainly concerned with the spatial and spectral resolutions, which relate to the scope, although there are no related elements in the range component. Spectral resolution is one of the most important concepts for the framework. It follows a top down approach method, where the concept is parceled into two subcomponents, i.e. the visible part and the infrared part. The visible is made up of three color segments, i.e. the RGB (red, blue, and green). The infrared part is also made up of three segments, i.e. thermal infrared, near infrared, and far infrared. The parameters suited for the slot are explicitly defined and include has_spatial_resolution, has_spectral_resolution, etc. [47] developed a simple ontological approach for remote sensing image classification. The prototype was built upon the expert remote sensing knowledge expressed in [54].

A. ONTOLOGICAL FRAMEWORK FOR OBJECT FEATURE EXTRACTION

After an image goes through a segmentation process, each region is characterised by a set of features. The feature extraction process from eCognition software follows the general upper ontology defined using the top down method [55]. The features are divided into six categories, namely LayerProperty, GeometryProperty, PositionProperty, TextureProperty, ClassProperty, and ThematicProperty. The selection of features of interest is performed by an expert to allow object detection. Figure 13 shows a hierarchical breakdown of object features from the six categories. GeometryProperty, TextureProperty, and ThematicProperty are important features in detecting forest objects [56].

B. ONTOLOGY MODEL OF THE LAND COVER CLASS HIERARCHY

The upper-level ontology is developed using concepts from land cover classification systems (LCCS) [36]. Figure 13 shows a hierarchically simplified way of representing classes of interest emanating from the main land cover class [53]. [55] designed an upper level ontology for the Chinese Geographic Condition Census Project [57]. Figure 14 depicts the design of an eight land cover ontology. The procedure was as follows:

1) The first step was to establish a set of important terms, in this case; Fields, Woodland, Grassland, Orchards, Bare land, Roads, Building and Water.
2) Classes and class hierarchies were then defined, A land cover class was defined through a top down approach.

1) ONTOLOGY MODEL OF THE DECISION TREE CLASSIFIER

Ontologies typically express two algorithms, namely decision trees and semantic rules [55]. [58], [59] used decision trees in the field of ontologies to cluster and classify image objects. Findings proved that decision trees enhance ontologies to granulate information, thereby increasing image classification accuracy. [59] uses decision trees to solve the problem of inconsistency between overlapping ontologies. [47] use decision trees for ontology matching; the matching process is purely based on derived decision tree rules for an ontology that are compared with rules for external ontologies. [55] designed an ontology model for decision tree classifier that

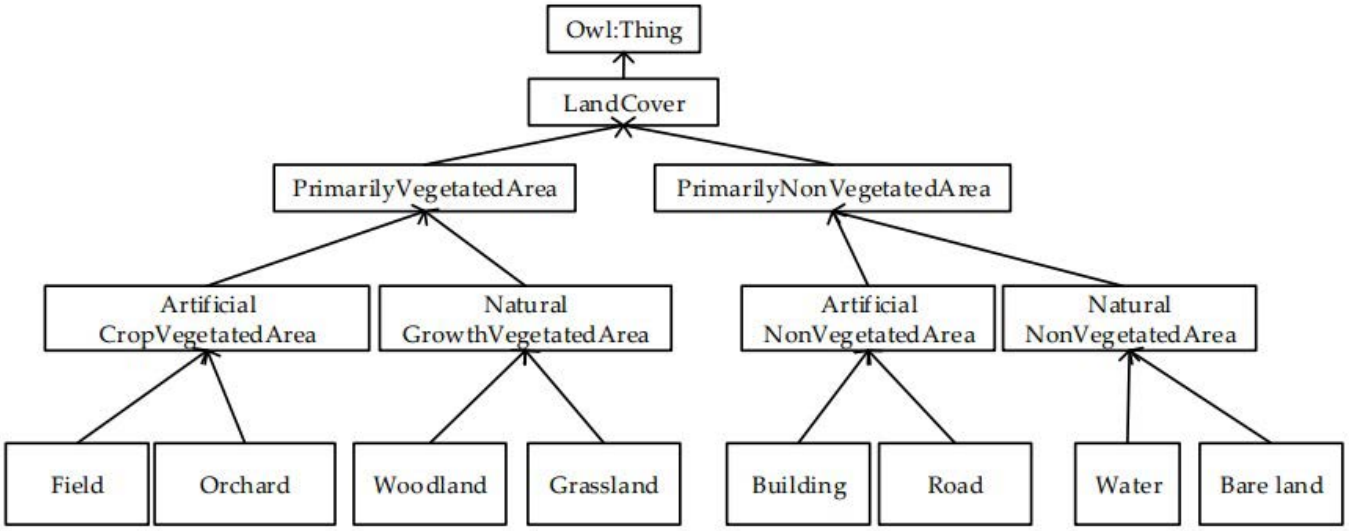


FIGURE 14. Land cover class hierarchy in ontology [55].

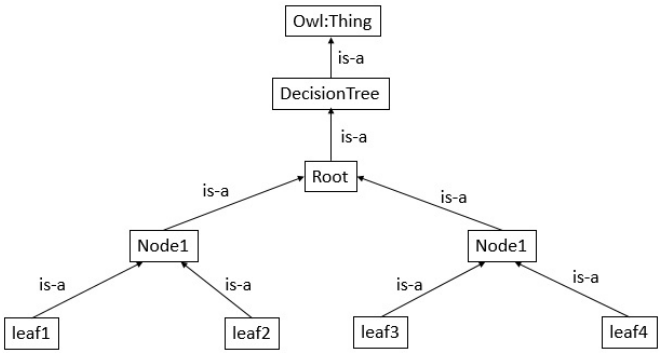


FIGURE 15. Ontology model of the decision tree classifier [55].

consists of three parts; (1) a set of decision trees is composed of all essential terms and concepts, for instance, a node and a leaf; (2) a slot is defined by the following inequality symbols $> \geq, <, \leq (3)$. The final step is to create the nodes. Figure 15 shows the elements of the decision tree classifier.

2) ONTOLOGY MODEL OF THE SEMANTIC RULES

[55] followed a two phased approach to designing an ontology model for semantic rules; the first is the establishment of mark rules, followed by decision rules. Mark's rules convert low level features to semantic concepts. On the other hand, decision rules are inferred from mark rules and apriori knowledge.

- Ontology model for mark rules

The morphology of semantic notions is classified into strip and planar; the shape is regular and irregular; the texture is smooth and rough; the brightness is light and dark; the height is high, medium, and low; and the position relationship is adjacent, disjunct, and contained. The ontology model of the mark rules is shown in Figure 16

- Ontology model of the decision rules

Ontologies explicitly represent concepts in the same way humans describe concepts in their domain of interest. However, ontologies that are developed disregarding decision rules have proved to be computationally expensive [60]. This is due to their inability to capture the kinds of decision-making knowledge that arises in practice, such as those involving multiple ontologies. Decision rules on ontologies help in three ways, namely: [61], [62]; (a) they take into cognisance primitives from multiple ontologies as well as primitives that are not part of the rule framework; (b) they are time dependant (c) they incorporate default assumptions. Eight types of land cover obtained from the Chinese Geographic Census Project [57] were defined in terms of a rule as outlined in Figure 17.

C. SEMANTIC NETWORK MODEL

Semantic networks graphically represent knowledge in the form of nodes and links, whereby links provide hierarchical relationships between objects [63]. The semantic network model explicitly express knowledge through concepts and their corresponding semantic relations [55]. This is shown in Figure 18. The network bridges the gap between low-level characteristics and high-level semantics, reducing the semantic gap.

D. ONTOLOGIES FOR KNOWLEDGE MANAGEMENT

Framework ontologies and domain ontologies are the two most important types of ontologies. Frameworks, or foundation ontologies, consist of concepts explicitly expressed in high-level knowledge (for human understandability), and they are also not designed for a specific domain. A domain ontology has knowledge tailor-made for a specific domain, e.g., remote sensing. Domain ontology eave drops from framework ontology. Domain ontologies have a hierarchical structure of two levels; the first level is called the ABox,

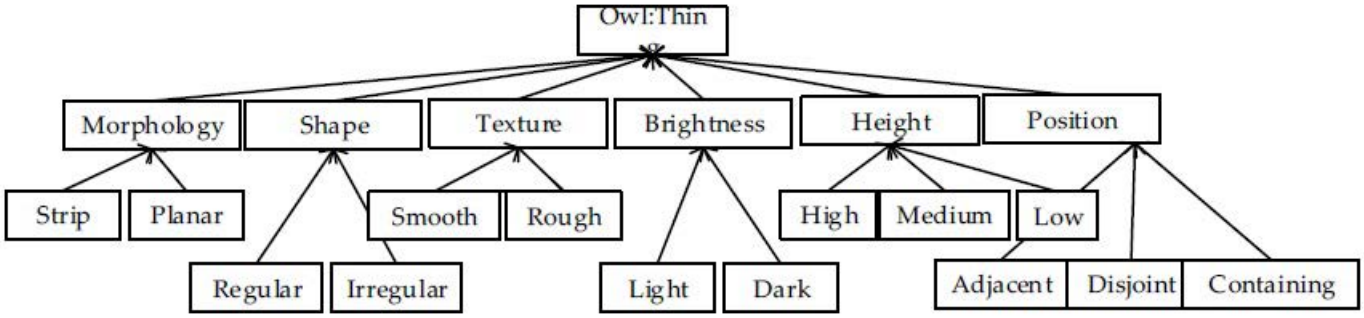


FIGURE 16. The mark rules in ontology [55].

1) Fieldland = Regular ∩ Planar ∩ Smooth ∩ Dark ∩ Low ∩ Adjacent to Road ;
2) Woodland = Irregular ∩ Planar ∩ High ∩ Rough ∩ Dark ∩ Adjacent to Fieldland;
3) Orchardland = Regular ∩ Smooth ∩ Planar ∩ Dark ∩ Adjacent to Fieldland;
4) Grassland= Irregular ∩ Planar ∩ Smooth ∩ Dark ∩ Low ∩ Adjacent to Building;
5) Building = Regular ∩ Planar ∩ Rough ∩ High ∩ Light∩ Adjacent to Road;
6) Road= Regular ∩ Strip ∩ Smooth ∩ Light ∩ Low;
7) Bareland = Irregular ∩ Planar ∩ Rough ∩ Lght ∩ Low;
8) Water =Irregular∩ Planar ∩ Smooth ∩ Dark ∩ Low ∩ Normal Differential Water Index(NDWI).

FIGURE 17. Decision rules based from ontology [55].

and the second level is called the TBox. ABox contains assertions (or rules) that comprise the theory that the ontology describes in its domain of application [64]. TBox is where experts conceptualise their knowledge in a specific scientific domain [47]. There are vast paradigms for modelling ontologies, but chief amongst them are Description Logics(DL) [65] and rule formalism. The DL formalism serves as a foundation for building ontologies using the web ontology language (OWL) [66]. Ontologies can be inferred from new knowledge using DL, which makes ontologies machine understandable.

E. MODULAR ONTOLOGICAL APPROACH

The modular approach is the best way of building complex ontologies from simpler (modular) ontologies in a constant and well-defined way [67]. Such an approach allows collaborative development by many different domain experts to build a single ontology through the integration of independently developed ontologies. The ontological approach is carried out in such a way that TBoxTs are not changed when elements of T' are reused in another TBoxT. Formalisation of such a property follows the conservative theorem [68].

Definition 1 (Conservative Extension): Let T and T' be TBoxes, Sig(α) be a signature of axiom α and Sig(T') be

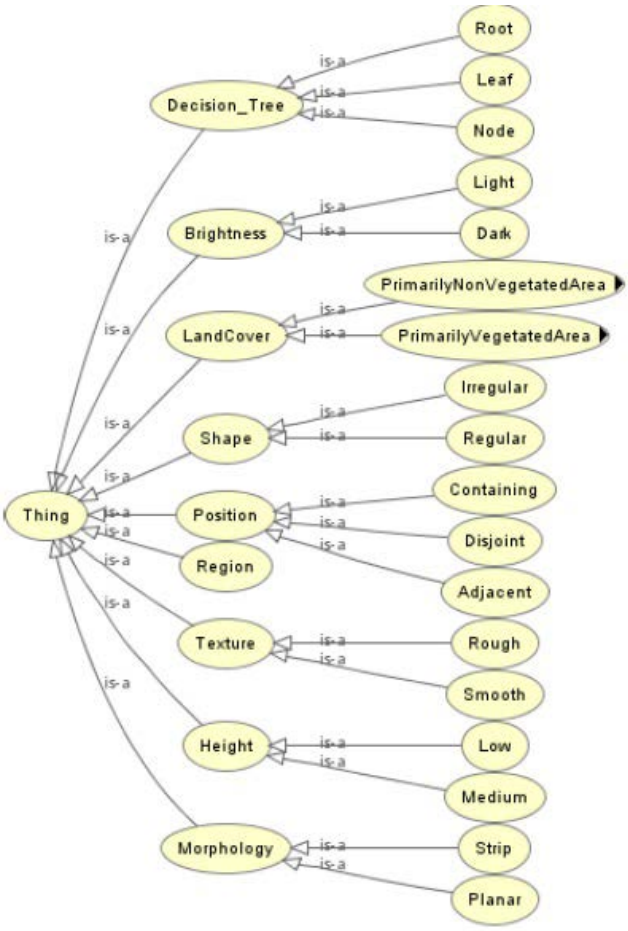


FIGURE 18. The semantic network model [55].

signature of TBoxes of T'. Then $T \cup T'$ is a conservative extension of T' if for every axiom α with Sig(α) $\subseteq$ Sig(T') we have $T \cup T' \Rightarrow \alpha$ iff T' $\Rightarrow \alpha$ [67]. In addition, if two independent parts T_1 and T_2 of an ontology T, are constructed in a modular way, then T remains modular as well. These are formalised as follows [67]:

Definition 2 (Modularity): Let Loc(T) be a local signature T and Ext(T) be external signature. A set M of TBoxes T with Sig(T) = Loc(T) $\uplus$ Ext(T) is a modular class if the following condition holds:

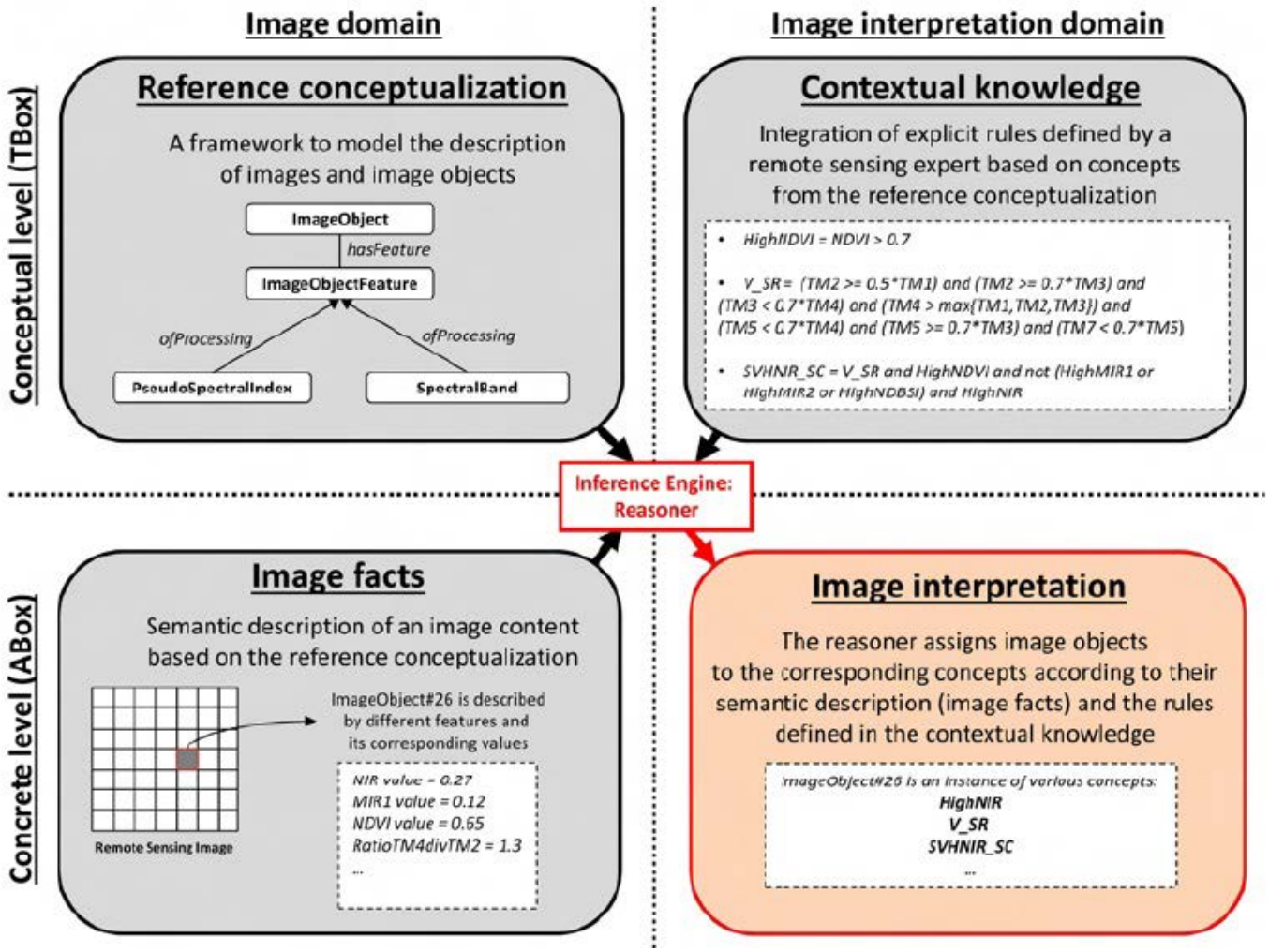


FIGURE 19. Structure of knowledge base [47].

M1. if T ∈ M, then T ∪ T' is a conservative extension of every T' such that Sig(T') ∩ Loc(T) = θ

M2. If T_1,T_2 ∈ M, then T = T_1 ∪ T_2 ∈ M with Loc(T) = Loc(T_1) ∪ Loc(T_2)

Falomir at al [66] proposed three levels of knowledge that are imperative for designing a modular ontological approach: the reference conceptualisation (which provides a description of images and image objects), the contextual knowledge (a set of rules defined by a domain expert) and the image facts (these are semantic descriptions of image content). Figure 19 illustrates how the reasoner assigns image objects to their corresponding concepts based on facts drawn from reference conceptualisation and contextual knowledge also drawn from reference conceptualisation.

(a) The reference conceptualisation

It is a general model for describing image objects in remote sensing. It consists of two packages, namely, (1) the image structure package and (2) the image processing package [47]. The image structure package is superimposed with the ImageObjects concepts, which describe objects according to their characteristics, and the ImageObjectFeature concept, which links related concepts with associations such as ''hasfeature''. The image processing package is composed of the PseudoSpectrallndex and SpectralBand concepts. The concepts help remote sensing experts describe contextual knowledge. Concepts such as spectral bands and texture are used by remote sensing experts to interpret remote sensing images.

(b) The Contextual Knowledge

Contextual knowledge's purpose is to represent remote sensing expert knowledge using DL, hence the name ''contextual knowledge.'' The basis of this knowledge comes from the Remote Sensing Science expert. As a result, it is a ''subjective'' description of image rules rather than an ''objective'' depiction of image structure. Figure 20 shows the concepts, relations, and instances in conceptual knowledge.

(c) The Image Facts

These are facts extracted from image analysis, and they are stored in the ABox [47]. The TBox contains the reference conceptualisation and the contextual knowledge [47]. Facts in ABox provide semantic descriptions of image objects, and

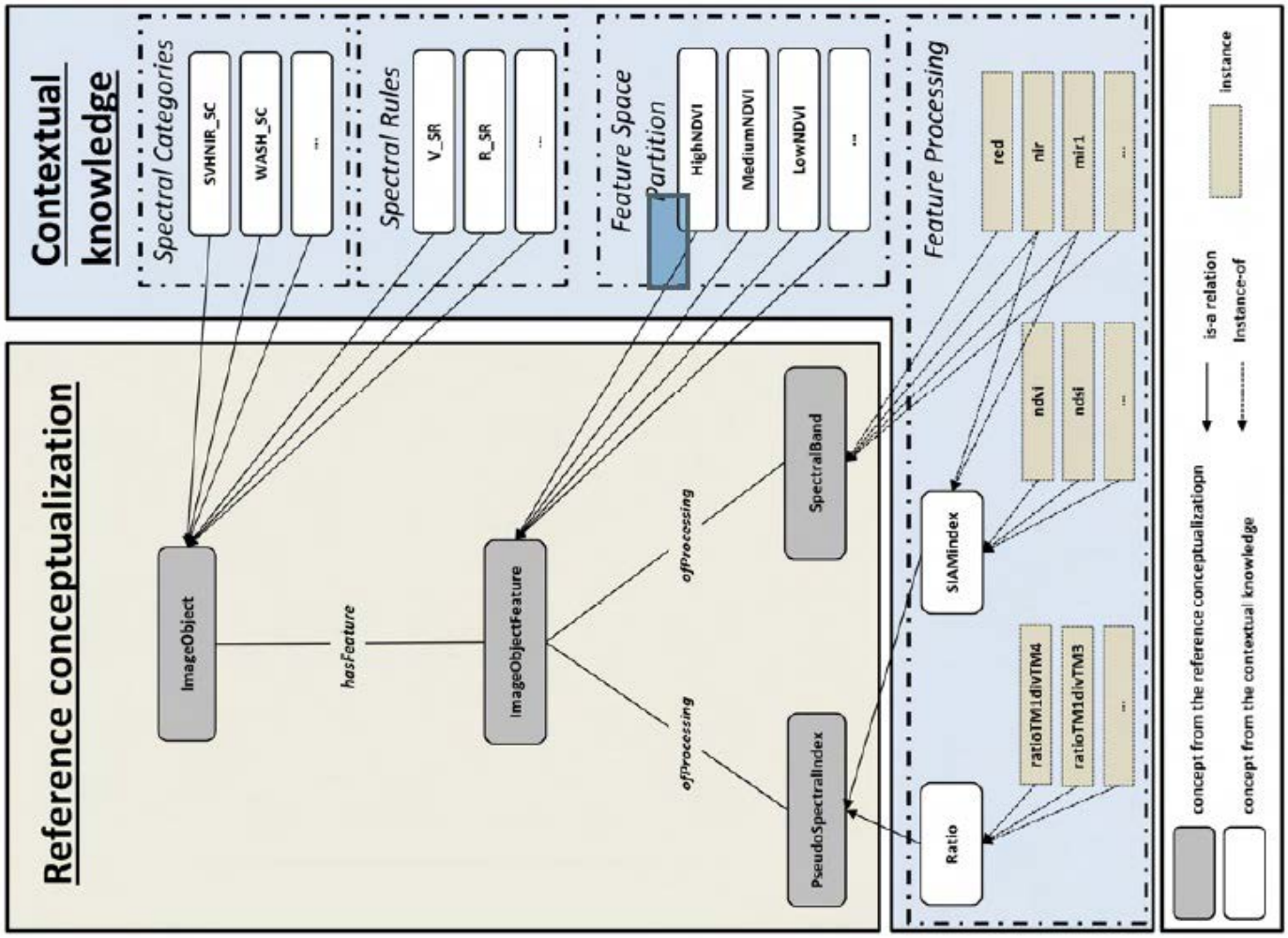


FIGURE 20. Conceptual knowledge showing concepts, relations and instances in an ontology [47].

the description is done with the help of reference conceptualisation and conceptual knowledge.

V. VEGETATION DETECTION

Unsupervised and supervised classification algorithms are very crucial in identifying vegetation areas.

A. UNSUPERVISED CLASSIFICATION INDICES

Spectral indices are used in these methods to detect vegetation areas. The Normalized Difference Vegetation Index (NDVI), which is calculated for each pixel in an image, is one of the indices utilized. The NDVI image is represented in a gray scale image. As shown in Figure 21: image (a) is a representation of an image using the RGB channel; image (b) is the representation of the same image in an NDVI format using the gray scale.

$$NDVI = \frac{\psi IR - \psi R}{\psi IR + \psi R} \tag{11}$$

Equation 4 illustrates the calculation of NDVI, where ψIR and ψR are pixel values in the infrared and the red band respectively. The formula defines vegetation as areas that have a higher reflective index in the infrared than the red band

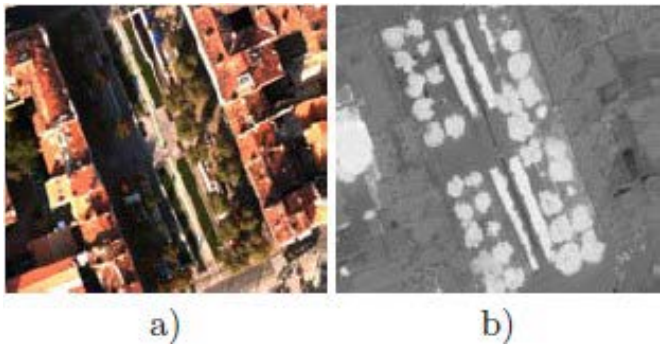


FIGURE 21. (a)RGB input image (b) NDVI image [69].

index. The formula was then refined to take into account the spectral index [70].

$$SI = \frac{\psi R - \psi B}{\psi R + \psi B} \tag{12}$$

Equation 5 illustrates the calculation of SI, where $\psi\ B$ is the pixel value in the blue band and $\psi\ R$ is the pixel value in the red band. An NDVI value and a SI value are binarized to create a vegetation mask. This is shown in Figure 22.

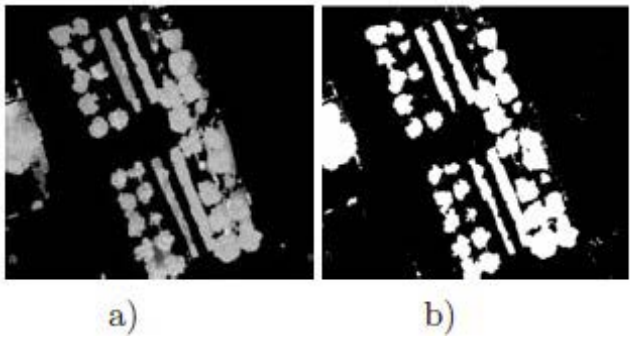


FIGURE 22. Image (a) SI image (b) vegetation mask obtained with the NDVI and SI indexes [69].

B. SUPERVISED CLASSIFICATION INDICES

Detection of vegetation by spectral indices is highly dependent on spectral characteristics. In other cases, supervised classification methods are primarily based on Support Vector Machines (SVMs). The feature vector that defines all pixels in the training data set contains four characteristics, namely: the reflectance value of each pixel in the infrared, red, green, and blue. Supervised methods do well in distinguishing between non-vegetation and vegetation areas through spectral indices. It necessitates the use of a SVM capable of determining the best linear separator. Random Forest (RF), k-Nearest Neighbour (kNN), SVM and sparse representations are among pixel wise classifiers that have been used for the last decade [71]. These traditional methods only consider spectral information as the basis of the classification process, disregarding spatial contextual information which contributes significantly to the classification performance [71]. Several researchers have proposed a hybrid of spectral-spatial classification that takes into account both the spatial context and spectral information, based on the assumption that pixels from a local region have similar spectral information. [71] proposed a hybrid model of kNN combined with guided filter for hyper-spectral image (HSI) classification of forest trees. Joint hybrid model of kNN and guided filter (PGF-kNN) was used to optimise hyper-spectral images produced by kNN. Optimised hyper-spectral images were taken in as input by the Joint kNN, and processed to produce the classification maps. Each class map was converted into a probability value and the class map with the highest probability value was chosen as the classification result. [72] conducted a study to determine the reliability of RF and SVM algorithms in the classification of very high resolution images (VHR), obtained from oak woodlands of a Mediterranean ecosystem. The first stage was data acquisition, where images were subjected to a Structure-Form-Motion (SFM) technique to identify common features in overlapping images. Each image was then orthorectified through the interpolated digital surface mode (DSM). Finally, all the images were combined into an orthomosaic. The workflow of the study followed 4 main steps, namely, preprocessing, segmentation, classification and accuracy assessment. Figure 23 shows the workflow of the proposed model. In the preprocessing stage each input layer was subjected

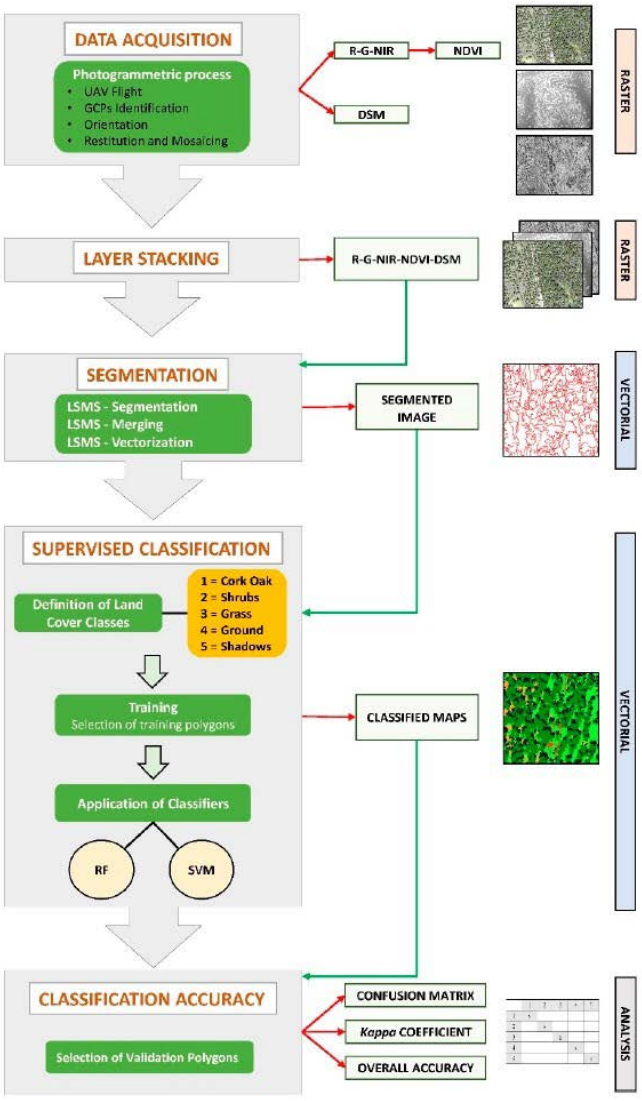


FIGURE 23. Workflow that presents the stages of preprocessing, segmentation, classification and accuracy assessment [72].

to a linear band covering a range of 8 bits, that is, from a minimum of 0 to a maximum of 255. The process was done to normalise each band, to suppress the effect of possible outliers on the segmentation. A layer stretching process was performed on images containing R-G-NIR (Red, Green, Near-Infrared) bands, obtained during spring and summer seasons through integrating NDVI and DSM data, to obtain the final 2 five band orthomosaics. Such a process was of significant importance because OTB segmentation requires only one rasta image as the input data. Spectral separability is of significant importance when it comes to image classification. The M-static defined in Equation 13 was employed [72] to measure the separability of NDVI and DSM layers of varying types of vegetation.

$$M = \frac{|\mu_1 - \mu_2|}{\sigma_1 + \sigma_2} \tag{13}$$

where, μ_1 is the mean value of class 1 and μ_2 is the mean value of class 2. σ_1 is the standard deviation of class 1 and

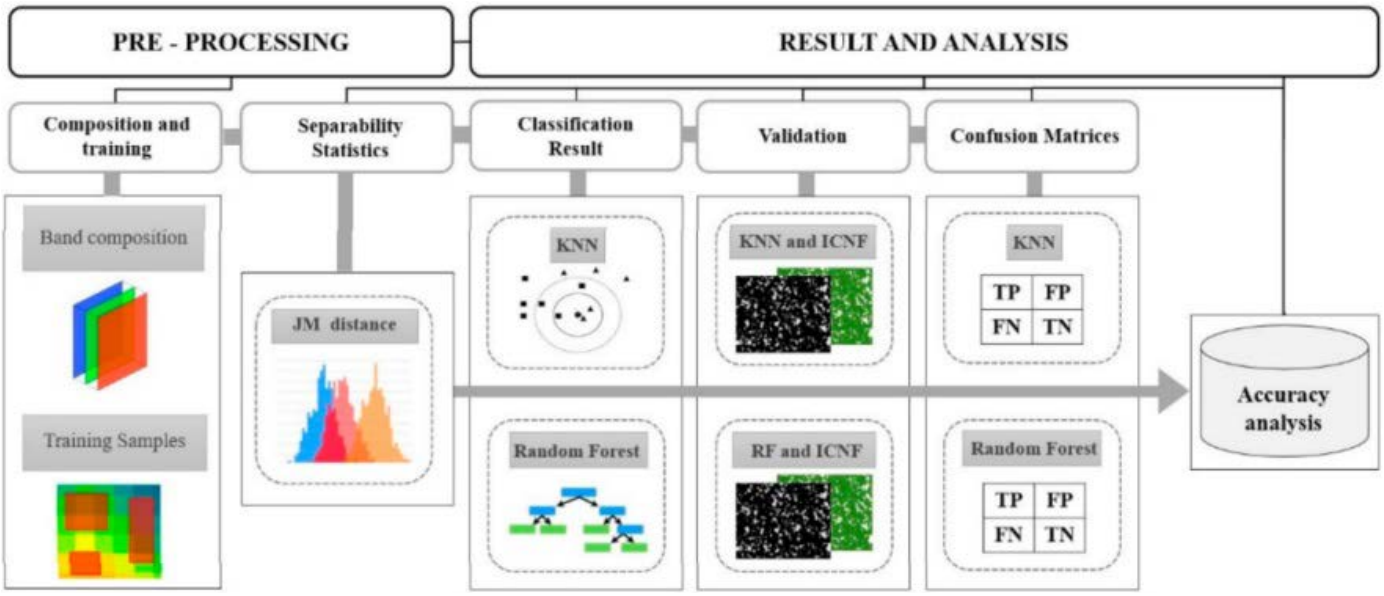


FIGURE 24. Flowchart of the model that harmonises RF and kNN [73].

σ_2 is the standard deviation of class 2. If M $<$ 1 it signifies overlap of classes, if M $>$ 1 it denotes that classes are well separable. The segmentation process considered both semantic properties and radiometric information. Large-scale mean shift (LSMS) segmentation was used in the study because of its ability to perform tile-wise segmentation of large VHR imagery [72]. The OTB LSMS segmentation process followed the steps of LSMS smoothing, LSMS segmentation, LSMS merging and LSMS vectorisation. Classification was performed for five different land cover classes, namely, grass, cork oak, soil, shrubs and shadows. Two supervised learning algorithms including RF and SVM were used to perform the classification. SVM performs linear separation in a hyperspace using a $\mu(.)$ mapping function. In the case where objects are not linearly separable, the kernel method is used where it takes into account projections of feature space [72]. RF uses decision trees for bagging to produce different subsets of variety of trees. Every decision tree in the RF participates in the classification process and the classification label returned is the class with the most votes.

Another study [73] analysed the performance of kNN and RF classifiers for mapping forest fire areas. The authors [73] implemented kNN and RF to classify forest areas and explained the effects of different satellite images on both classifiers. Figure 24 shows the flow chart of the model. The model being a supervised approach was implemented by using multi-spectral images obtained from Landsat8, Landsat-2, and Terra sensors. The classification accuracy was determined by the confusion matrices. The machine learning classifier based on kNN and RF produced excellent results with k set to 5 for kNN and 400 trees for RF. The results from the hybrid model achieved a very high classification accuracy with an Overall accuracy (OA) $>$ 89% and Dice coefficient (DC) $>$ 0.8. Other studies [74], [75] have also implemented non-parametric algorithms such as kNN and RF in remote sensing applications.

VI. IMAGE SEGMENTATION

An input image is partitioned (or subdivided) into meaningful image objects (segments). Image segmentation can be classified into two categories: supervised (empirical discrepancy methods) and unsupervised (empirical goodness methods) [76]. Unsupervised approaches evaluate a segmentation result based on how well the image object matches a human perception of the desired set of segmented images, and they use quality criteria that are typically created in accordance with human perceptions of what constitutes a good segmentation. Supervised methods compare a result from segmentation with a ground truth [2]. If ground truth can be reliably established, supervised methods are preferred.

A. TYPES OF IMAGE SEGMENTATION

Pixel, edge, and region-based image segmentation methods are the three primary types of traditional image segmentation. [77].

(a) Pixel Based Methods

This method involves two important processes: (1) image thresholding and (2) segmentation in feature space. For image thresholding, image pixels are divided according to their intensity level [78]. There are three types of thresholding [79], [80]:

(1) Global thresholding - T being the appropriate threshold value. The output of an image q(x,y) based on T is obtained from an original image p(x,y) as

$$q(x, y) = \begin{cases} 1, & ifp\,(x, y) > T \\ 0, & ifp\,(x, y) \leq T \end{cases}$$

(2) Variable thresholding - This when the value of T varies varies over an image and it comes in two flavours:

- Local Threshold - T depends on the neighborhood of x and y.
- Adaptive Threshold - T's value is a function of x and y.

(3) Multiple thresholding - It has multiple values of T. The output image is computed as follows:

$$q(x,y) = \begin{cases} m, & ifp\,(x,y) > T1 \\ n, & ifp\,(x,y) \leq T1 \\ 0, & ifp\,(x,y) \leq T0 \end{cases}$$

However, these methods suffer from incomplete segmentation, so the output results need to be clumped. Also, these methods are appropriate for images with lighter objects than the background.

(b) Edge Segmentation methods

Edge-detecting operators are employed to detect all possible edges that are found in an image. Adjacent edges are clearly separated by a gray sharp edge, but there could be a case where the gray value is not continuous [81]. The edges will be represented by discontinuity in gray level, color, texture, etc. This discontinuity is detected by using derivative operations such as differential operators [82]. The Prewitt, Roberts, and Sobel operators are the most frequently utilized first order differential operators [83]. There are a number of edge detection operators such as the template matching edge detectors. One challenge with edge-based segmentation is that sometimes it presents edges in locations where there is no border. Filtering, enhancement, and detection are the three processes in edge segmentation algorithms [77]. The purpose of the filtering process is to reduce the amount of noise present in the imagery. The enhancement uses high pass filtering to detect and reveal local changes in intensity. Finally, the edges detected (using threshold techniques) are combined or linked together to form the boundaries of the image object. One challenge with edge-based segmentation is that, sometimes it presents edges in locations where there is no border.

(c) Region Based Segmentation

Images are segmented into regions with similar properties using region-based approaches [81]. There are three types of region based segmentation, namely: (1) region-growing segmentation, (2) region-splitting and merging segmentation [82], [84].

(1) Region Growing Segmentation

It starts with the matrix's origin (seed point), which is then subjected to a rule that joins surrounding pixels to these starting regions, and the procedure is repeated until a particular threshold is met [81]. The method is repeated until there are no more pixels to ascribe. This process is repeated until the entire image is segmented. The algorithms, on the other hand, suffer from a lack of control over the region's growth break-off criterion [85].

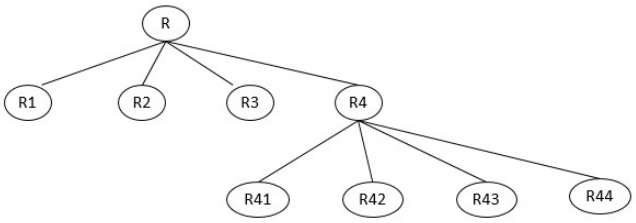


FIGURE 25. Region splitting.

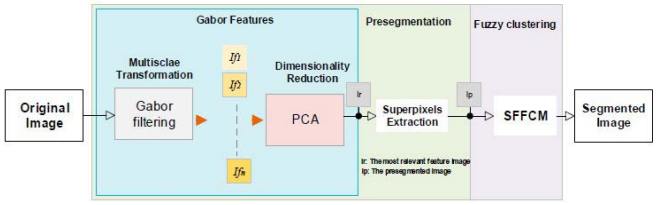


FIGURE 26. Image segmentation state of the art [76].

(2) Region splitting and region merging

The original image is split or subdivided into sub images. Each sub-image is recursively divided into its own sub-images based on the condition or predicate given. If the condition is not satisfied, further splitting ceases [82]. Figure 25 shows the splitting process.

B. IMAGE SEGMENTATION STATE OF THE ART

Reference [76] proposed a segmentation process that improves segmentation accuracy by modifying the super-pixel extraction methodology so as to increase robustness to added noise. The segmentation method is based on Gabor filtering and Principal Component Analysis (PCA). Figure 26 presents the state-of-the-art segmentation process. The method depends on two principal tasks: (1) pre-segmentation (super-pixel extraction), and (2) clustering of previously extracted pixels.

(a) Pre-segmentation

An input image is subdivided into a number of regions of interest. Each region is made up of pixels with similar features. The Watershed Transform (WT) clustering based super-pixel algorithm has previously been considered for super-pixel extraction [86], [87].

(b) Gabor filter

Gabor filters are used to extract spatially localized spectral features [76]. They have been advocated for because they are based on principles found in similar human visual systems and have key features that can be utilized to segment images.

Before the introduction of deep learning, machine learning techniques such as SVM, K-means clustering, Random Forest, etc., were the chief algorithms for image segmentation. Semantic segmentation using deep learning has proven to work better than the aforementioned techniques because they classify each pixel of an image rather than the entire

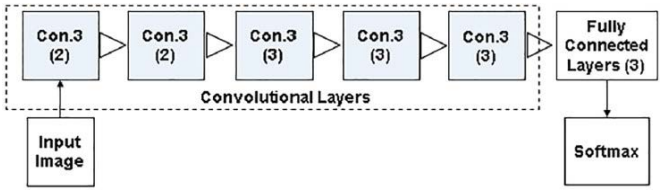


FIGURE 27. AlexNet [89].

image object. The next chapter gives an overview of semantic segmentation techniques.

VII. SEMANTIC SEGMENTATION USING DEEP LEARNING

This section introduces fundamental ideas of CNNs and subsequent variants for semantic segmentation, as well as their network structures [88].

A. AlexNet, VGGNet AND GoogleNet

These are the three chief deep neural networks for image classification, which formed the major foundations of later developments. The networks support network architectures for semantic segmentation.

1) AlexNet

AlexNet is made up of five convolutional layers and three connected layers [89]. In between the convolutional layers is a pooling layer whose role is aimed at reducing dimensionality and computational complexity. AlexNet's pooling strategy is max pooling, and the strategy is to obtain the biggest value covered by the filter, which is used to remove noisy components [88]. Filters of sizes 11×11 and 5×5 are used in the first and second convolutional layers, respectively. The last three layers use small-sized filters of 3×3. The whole process is described in Figure 27. The primary purpose of such filters is to be solely used for feature extraction. Varying filters accommodate objects of different scales.

1) It supports the application of non-saturating Rectified Linear Unit (ReLU) whose output is defined by

$$F(x) = max(x, 0).$$

2) It employs the overlapping max pooling strategy (which means that each filtering operation's step size (stride) is smaller than the filter's overall size).
3) To reduce over-fitting, it uses the dropout approach in fully-connected layers.

2) VGGNet

The network is made up of three fully connected layers and a varying number of convolutional layers. This is shown in Figure 28. Unlike AlexNet, VGGNet has fixed small size filters of 3×3 in the convolutional layer [90]. The number of weights in the network is reduced by using small filters, which minimizes the training complexity. Just like AlexNet, VGGNet uses max pooling over a 2×2 window slide of 2 pixels. The advantage of simplifying convolutional layers

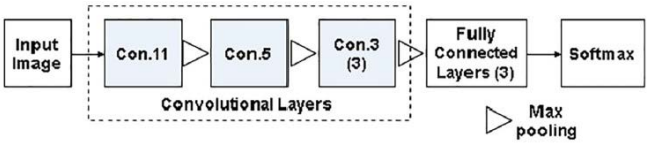


FIGURE 28. VGGNet [89].

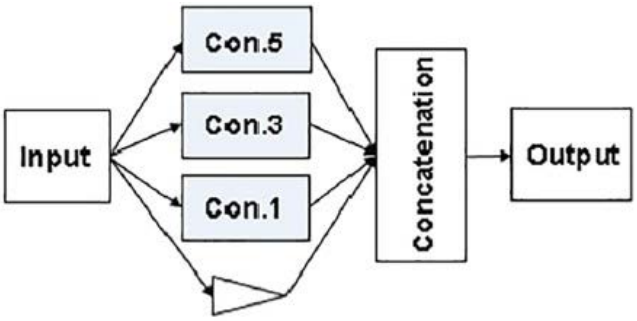


FIGURE 29. GoogLeNet [89].

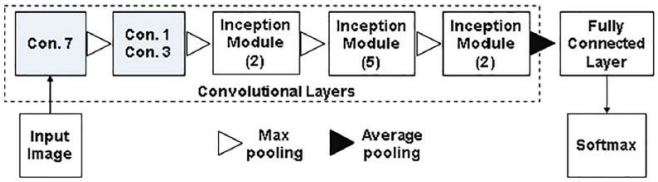


FIGURE 30. Inception Module [69].

to a greater extent is that it increases network depth, thereby improving the accuracy of the network. The network's performance in tasks like semantic segmentation and target detection is improved by using features extracted from CNN that are structured in a hierarchy of scales [91]. Other classifiers, such as SVMs, can use the features without fine-tuning [92].

3) GoogLeNet

The architecture is different from the other three in that it involves three aspects, namely the inception module, at the training stage, an auxiliary classifier is required, as well as one fully connected layer [93]. Output results from these filters are concatenated with the maximum pooling result. Between the inception modules, maximum pooling is employed, and after the last inception module, average pooling that employs dropout is used [94]. The flow chart diagram is shown in Figure 29. The network is so deep because it is made up of nine inception modules and up to three convolutional layers. Because of the profundity of the network, the smooth flow of gradient from layer to layer becomes an issue. Figure 30 shows the Inception Module. The issue is addressed by adding an auxiliary classifier in the middle of the convolutional layers, whose role is to process the outputs from the inception modules. The loss from these classifiers is added to the overall loss of the network during training. Auxiliary classifiers are prohibited from making decisions during the prediction phase.

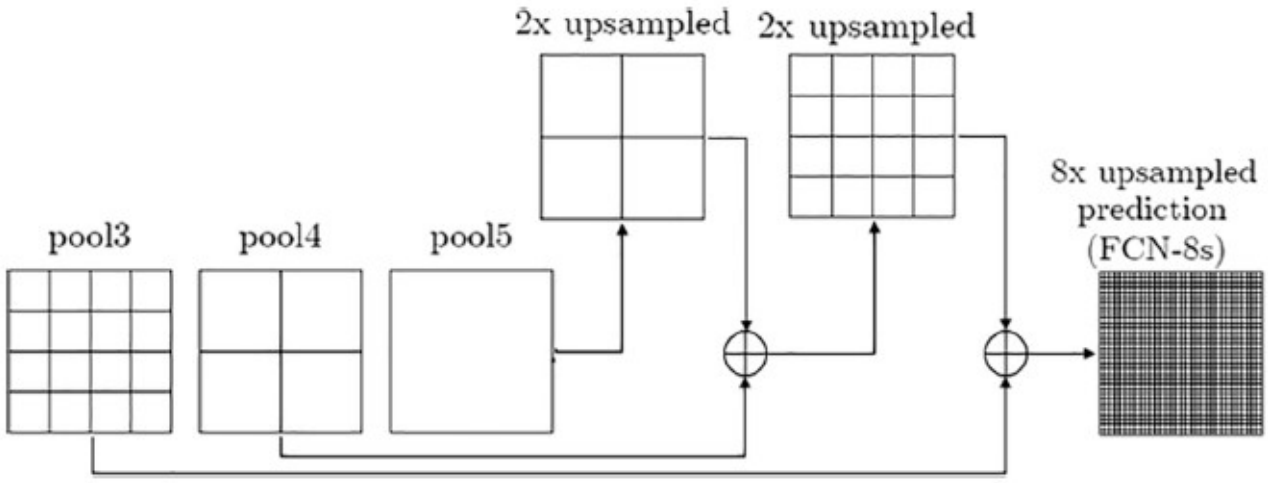


FIGURE 31. FCN Network [95].

B. FULLY CONVOLUTIONAL NETWORK

Fully convolutional networks for semantic image segmentation are an extension of AlexNet, VGGNet, and GoogLeNet [95]. Multi-convolutional, deconvolutional, and fusion are the three steps that define the network. The flow chart is shown in Figure 31. Convolutional layers have been substituted for fully linked layers, with the specification that each image's score be computed using a 1×1 convolution. Because of pooling, the output image from convolutional layers is smaller than the input image. The deconvolutional process is used to restore the image. It uses the same methods as the convolutional process, but cushions the framework (by padding the matrix) and joins the elements inside a deconvolution filter to increase the input size. The process of recovering the original image through the deconvolution process has some side effects; for example, some details are lost as a result of the dilution of class scores. To circumvent the side effects, the skip architecture combines semantic information obtained from layers with location details obtained from previous layers. By element wise addition, the up-sampled deep layer is fused with the yield or output of a shallow layer.

C. UNET

The building blocks of Unet are the convolutional and deconvolutional layers. The network works well with small images, hence the paramount step is downsizing of input images [96]. Convolutional layers use filters of size 3×3 which produce output images that are subsequently subjected to Relu for processing, followed by maxpooling (which uses a stride of two). Maxpooling generates downsized outputs. Feature channels in the convolutional layers double at each and every step. The deconvolutional layer does upsampling, but a 2×2 convolution is used to limit the number of features to the required standard. The network generates the segmentation result by applying a 1×1 convolution on the feature map and labeling pixels. The interconnection of layers in Unet is shown in Figure 32.

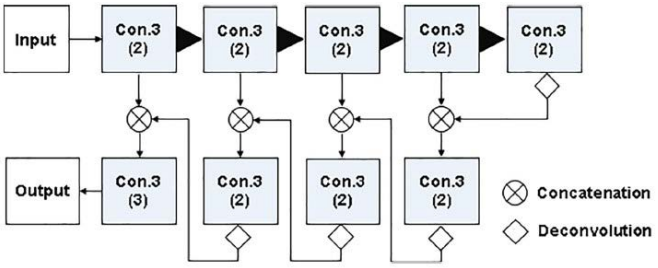


FIGURE 32. Unet [89].

D. SegNet

The network is composed of two subnetworks, namely; the encoder and decoder networks. The encoder network's mandate is the downsizing of feature maps. It consists of a varying number of convolutions and subsequent maxpooling operations for feature extraction [97]. However, features produced have vague or ambiguous spatial information. The issue is solved by saving an element index that will be used later in the decoder network's up-sampling procedure. Convolutions map low-resolution features to high-resolution features in the decoder network. A 2×2 low-resolution feature, for example, is up-sampled to a 4×4 matrix. This process may result in the loss of spatial information; therefore, reusing the pooling index from the encoder network completely recovers the lost information. The SegNet network is depicted in Figure 33.

E. DeepNet

It is a variant of FCN that employs dilated convolution to broaden the scope of filters to include image context in a larger neighborhood while also allowing for flexibility over feature response resolution [17]. Deeplab uses Atrous

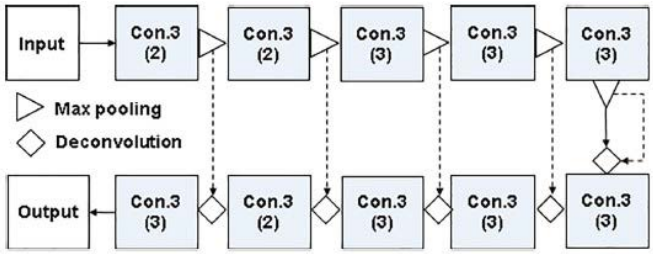


FIGURE 33. SegNe [95].

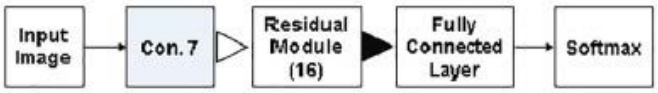


FIGURE 34. Residual Net [89].

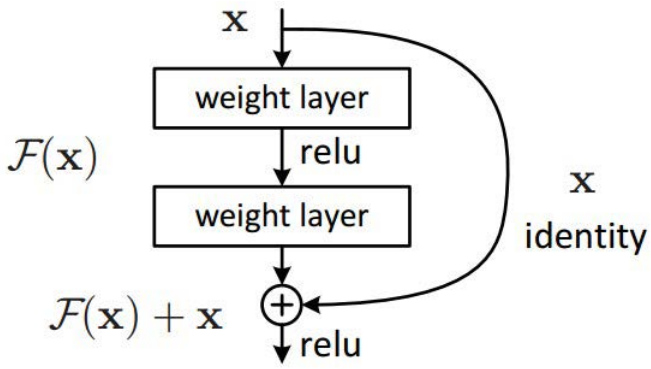


FIGURE 35. Residual block [95].

Spatial Pyramid Pooling (ASPP) for up-sampling. Several atrous convolutions operated on the same kernel but with various sampling rates are used in the scheme. An additional operator combines the output from all convolutions. Down-sampling processes and subsequent maxpooling operators make segmentation results lose some fine details. To solve the problem, conditional random filters (CRFs) are employed to improve the spatial localization of segmentation. CRF models contribute to the smooth segmentation process based on the underlying image intensities [98]. They boost the accuracy score by 1% to 2%.

F. ResNet

The residual network is well recognized for its 152 layer depth and residual block introduction [99]. The residual block is presented in Figure 35. As based on traditional neural networks, the greater the number of layers, the better the performance of the network. However, because of the vanishing gradient problem, first layer weights will not be updated correctly through the backpropagation algorithm [100]. As the error gradient is propagated to earlier layers it goes through a repeated multiplication process such that the gradient becomes very small hence the network performance gets saturated and will start to decrease. This problem is solved by using the identity function, whereby the gradient is multiplied by one so as to preserve the input and avoid any loss in the information. The network is made up of the following components; 3×3 filters, CNN downsampling layers with a stride of 2, global average pooling, and a 1000-way fully connected layer with softmax at the end. ResNet employs a skip relation, which means that an original input is also connected to the convolution block's output. This aids in the solution of the vanishing gradient problem by allowing the gradient to flow in a different direction. The network diagram of the residual network is shown in Figure 34.

G. APPLICATION OF DEEP LEARNING TECHNIQUES

New emerging technologies such as deep learning have gained ground in the remote sensing science fraternity because the automatic processing of images by these techniques chiefly depends on human expert knowledge, which has impacted the way land surveys are done [101]. The main advantage of deep learning approaches is the automatic computational extraction of features, unlike other machine learning algorithms where feature extraction is typically manual [102]. The strength of deep learning algorithms lies in learning from examples. The learning process consists of a number of steps: first, an architecture of a network of nodes is clearly defined. The nodes that form an Artificial Neural Network (ANN) are arranged into layers. An ANN with many layers is referred to as a Deep Neural Network (DNN). The behaviour of the DNN is determined by the type and number of nodes as well as the connection between the nodes [101]. If an existing DNN is to be customized for an new application context, its weights are recursively updated to achieve the new desired response. This process is referred to as "transfer learning". Deep learning was originally used for locating and classifying different tree species in a mosaic built from UAV-acquired images [103], [104]. [105] devised a deep learning technique to detect and identify tree species. The objective of the study was to classify patches corresponding to tree species. The authors developed a Deep Learning (DL) architecture, which is a hybrid of ResNet and UNet, to come up with a semantic segmentation algorithm for tree spices that is precise and efficient. Seven orthomosaic images were collected using UAV in the winter, and one orthomosaic image was collected using UAV in the summer. The algorithm pipeline is presented in Figure 36. The first step of the technique identified the classes corresponding to each mosaic patch. The focus was on classifying the pixels in each mosaic patch. The incorporation of the ResNet architecture into the DL network enhanced the accuracy and efficiency in classifying forest images [104], [106]. Images were divided into patches in response to the prescribed annotations, and each patch was assigned to a list corresponding to the classes that matched it. Patches could belong to more than one class, resulting in patches having to be labelled repeatedly. Because of the repeated labelling of patches, the algorithm is referred to as a Multi-label Patch (MLP) based classifier. The ResNet architecture went through the training

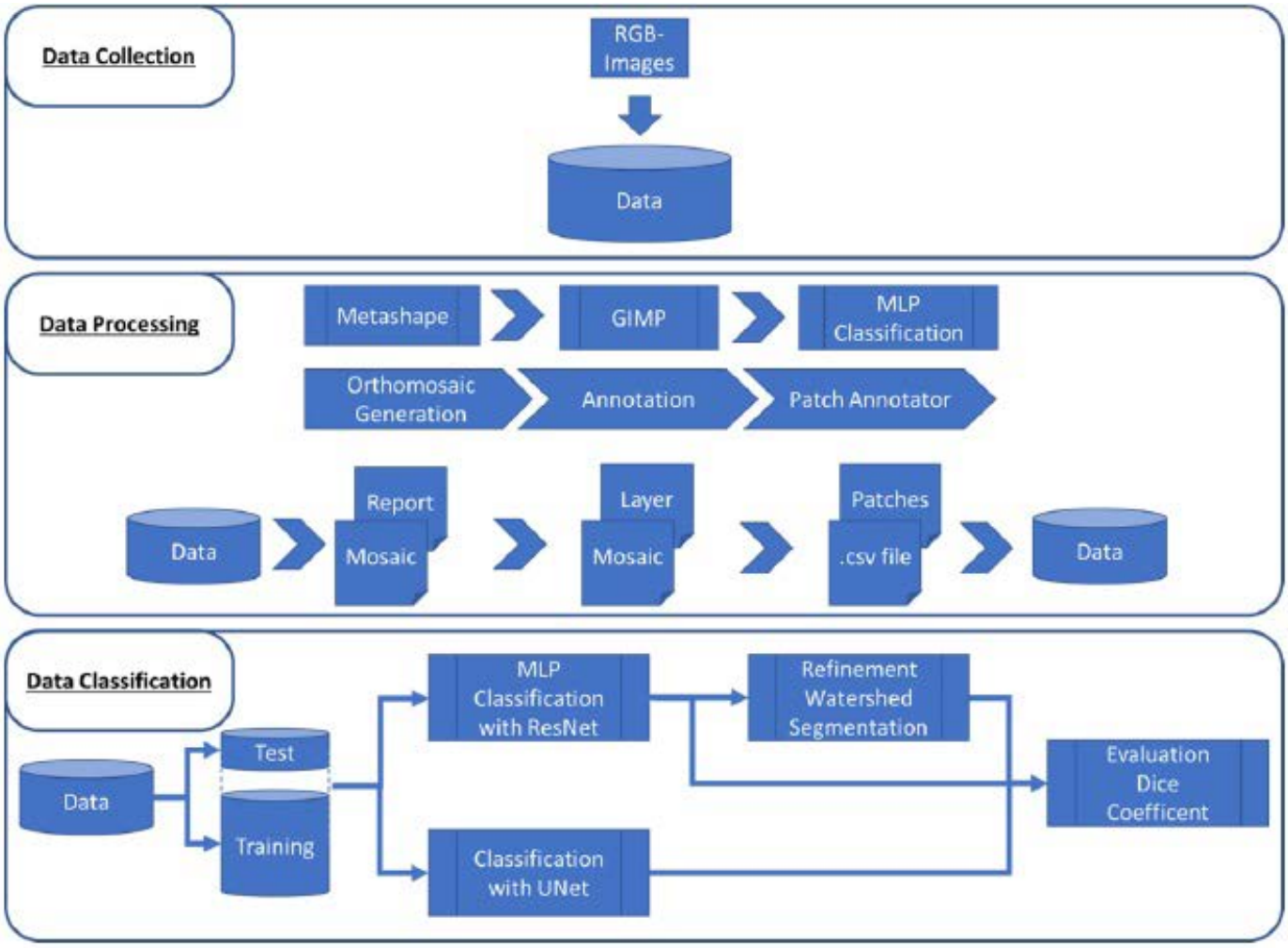


FIGURE 36. Algorithm pipeline [106].

phase so that it would be able to classify the patches. The MLP classification algorithm produced coarsely segmented images. A watershed segmentation algorithm was applied to refine the segmentation process. The UNet architecture, originally used for medical image segmentation [96], is also very useful for remote sensing images. The UNet architecture was trained with data and pixel-wise annotation patches. The segmentation process follows a number of steps: (1) mosaic images were split into patches for processing, (2) a UNet model was trained to predict patch segmentation, and (3) patch joining was used to obtain semantic segmentation for the entire mosaic image. The model achieved an effective learning transfer with a 12.48% improvement over random weights. Overall, the model reached a higher accuracy of nearly 95%.

Another study [104] proposed a Residual Neural Network (ResNet) architecture for classifying tree species acquired using a camera mounted on a UAV platform. In temperate forests, UAV images have been successfully used to distinguish between living and dead forest species [107]. The motivation of the study was that, most of the existing methods for tree species classification are cost-sensitive because they require very large data sets and are restricted to specific tree species [108]. The study proposed a model based on CNN to classify tree species at an individual level by analysing high resolution RGB images obtained from the UAV. A CNN was chosen in the study because of its ability to learn highly descriptive features from tree canopies. The study proposed a CNN model with 50 convolutional layers, referred to as ResNet50. Figure 38 shows the architecture of ResNet50. The procedure for performing tree crown delineation was based on the iterative local maxima filtering technique that was used to identify probable tree tops. Tree tops were designed as markers, hence a marker controlled watershed segmentation was performed as a means of complementing the DSM for segmenting the tree crowns. Figure 37 shows a tree crown segmented polygon. The tree crown delineation process enables tree crown identification labelling. In the training phase, images were shuffled in unison with their corresponding labels to randomise the input data so that the neural network becomes generalised. The model achieved an overall classification accuracy of 80%. The study concluded

FIGURE 37. Tree crown delineation [104].

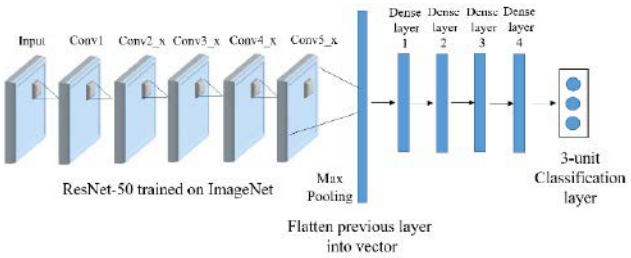


FIGURE 38. CNN model architecture [104].

that classification accuracy increases with an increase in the number of training images.

The task of classifying and mapping vegetation images has been difficult because the conventional methods employed are highly labour intensive. Deep learning and CNN came as solutions to the problems posed by traditional methods, but they are still not efficient in detecting ambiguous objects [109]. There is a little research that employs CNN to detect and classify vegetation in remote sensing science images [109]. A study by Guirado [110] successfully used CNN to detect wild shrubs from Google Earth images. The author demonstrated that a CNN is much better than traditional object detection methods. Another study [109] used a deep learning model and the chopped picture method to detect vegetation from Google Earth images. The study was carried out against the backdrop that existing work still faces huge challenges in classifying vegetation that has ambiguous and amorphous shapes, such as clonal plants. The training data was prepared using the chopped picture method, and images were put into two sets; one set with images completely covered with bamboo trees and the other set without bamboo trees. Images were then chopped into small squares and subsequently used as training images. A classical deep learning model in the form of a LeNet network was employed by the study because it is efficient in processing small-sized images. The network is composed of two convolution layers, two pooling layers, and one fully connected layer. The final layer was used to detect bamboo coverage in Google Earth images. Input images were randomly shuffled to alleviate overlapped training and validation data. 72% percent of the data was used for training and 25% of the data for testing. The model achieved an average classification accuracy of 97.52%.

VIII. FEATURE EXTRACTION TECHNIQUES

This section delves into the main techniques for feature extraction, and these include (1) Principal Component Analysis (PCA); (2) Independent Component Analysis (ICA); (3) Linear Discriminant Analysis (LDA); and (4) t-Distributed Stochastic Neighbor Embedding (t-SNE).

1) PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA is popularly used as a dimensionality reduction technique [111]. It was first proposed by [54]. From the original data input, the PCA method tries combinations of input features in order to determine the best features that summarise the original data. This is accomplished by looking at pair-wise distances to maximize variances and minimize reconstruction error [112]. Since PCA is an unsupervised learning algorithm it leads to misclassification of data in some cases [111]. Distortion errors arise when data is reconstructed back because samples would have been projected onto a subspace [113].

2) INDEPENDENT COMPONENT ANALYSIS (ICA)

ICA, like PCA, is a linear dimensionality reduction method that combines discrete components to produce input data with the goal of correctly identifying each of them [111]. It is based on the principle that two features are deemed independent if their linear and nonlinear dependence are both zero [114]. Independent Component Analyses are extensively used in medical applications such as Electroencephalography (EEG) and Functional Magnetic Resonance Imaging (FMRI) analysis to differentiate useful from unhelpful signals [111].

3) LINEAR DISCRIMINANT ANALYSIS (LDA)

LDA is a supervised learning dimensionality reduction technique and a machine learning classifier [111]. The method is similar to PCA in the sense that it calculates the projection of data along a direction, but instead of maximising variation of data, the LDA uses label information to determine a projection by maximising the ratio of between class variance to within class variance [113]. The goal of LDA is formulated as the Fisher criterion [115].

$$J(u) := \frac{u^T S_B U}{u^T S_W U} \tag{14}$$

Recently, this technique has been used for indoor positioning or localisation systems for the purpose of obtaining superior and higher accuracy [116]. The performance of LDA in the construction of data using independent variables is directly proportional to the number of data patterns [116]. However, its performance is yet to be confirmed in the context of non-linearity [117].

4) LOCALLY LINEAR EMBEDDING (LLE)

The LLE is built on a foundation of manifold learning. A manifold is a D-dimensional object that is embedded in a higher-dimensional space. A manifold is considered as an integration of small linear patches, which is done through piece-wise linear regression [118]. To do the integral operation, [119] proposed the construction of a kNN graph similar to an isomap. Then all the sample data is represented by a weighted summation of its k nearest neighbors. Considering w_i to be row i of the n x k weight matrix w, the solution to the goal is found by:

$$W_i = \frac{1}{1^T(G_i^{-1}T)} G_i^{-1} 1 \tag{15}$$

$$G := (x_i 1^T - V_i)^T (x_i 1^T - V_i) \tag{16}$$

where G is called a Gram matrix and V is a n x k matrix. After the process of representing samples as a weighted summation of their neighbors, LLE represents samples in the lower dimensional space by their neighbors with the same obtained weight.The method has been successfully used in feature extraction of Motor Imagery Electroencephalography (MI-EEG) and it outperformed methods such as Discrete Wavelet Transform (DWT) in classification accuracy with fewer feature dimension [120].

5) T-DisTribuTed STOCHASTIC NEIGHBOR EMBEDDING (T-SNE)

tSNE is an improvement of Stochastic Neighbor Embedding (SNE) [121], which is used for data visualisation. The main goal is to preserve the joint distribution of data samples in the original and embedding spaces. Considering P_{ij} and Q_{ij} to donate the probability that x_i and x_j and are neighbors and y_i and y_j are neighbors, it follows that:

$$p_{ij} = \frac{P_j|i + P_j|j}{2n} \tag{17}$$

$$p_{ij} = \frac{exp(-||x_i - x_j||_2^2/2\sigma_i^2}{\sum_{K\neq 1} exp(-||x_i - x_k||_2^2/2\sigma_i^2} \tag{18}$$

$$q_{ij} = (\frac{1 + ||y_i - y_j||_2^2)^-1}{\sum_{k\neq 1}(1 + ||y_k - y_l||_2^{2)^{-1}}} \tag{19}$$

Embedded samples are then obtained by adopting the gradient descent method over minimizing Keullback-Leibler divergence [122] of p and q distributions. The main advantage of t-SNE is the ability to deal with the problem of visualising ''crowded'' high dimensional data in a low dimensional space (e.g., 2D or 3D) [122], [123].

A. FEATURE EXTRACTION STATE OF THE ART

In image retrieval, calibration, classification, and clustering, it is critical to extract useful features or characteristics from the image [124]. Color histogram is the most significant method to represent color features [125]. [126] provided a state-of-the-art feature extraction model that consists of two parts: (a) adaptive color region extraction via the definition circle (DC) model, and (b) corner feature extraction via the edge detection model, which includes a suppression mechanism.

The purpose of the algorithm was to produce a clear and precise forest saliency map. The algorithm is broken down into three parts, and those are: (a) the color feature extraction part; (b) the determination of the center of the DC model; and (c) an accurate description of color. The algorithm is expressed in figure 36.

(A) Colour feature extraction

Model appropriate for the extraction of color features is the DC model, which is comprised of the following steps: (1) using the RGB picture G histogram to calculate the DC model's center; (2) mapping the image to the HIS color space or lab color; (3) using the k-means procedure to find the DC model's radius. The flow chart of the DC model is shown in Figure 41.

(1) Determine the center of the of the DC model

While the DC model can describe color fluctuations under specific gradients, the forest region's dominating hue is generally green, implying that the 'greenish' pixels in the forest area must be filtered off. As a result, the G channel (green) in the RGB three-channel system will be the focal point for filtering out pixels that fall within a given range and calculating the mean value within the range. That value will be regarded as the center of the circle.

(2) Color description

It is critical to note that the purity of the green is determined by the circle's center, thus the radius must be adjusted to account for a variety of color variations and fault tolerance. The RGB channel, on the other hand, does not function well for color adjustments. The RGB color system is converted to Hue, Saturation, and Intensity (HSI) or lab color space to fix the problem. The color can be defined more correctly using only two channels, namely hue and saturation, rather than the RGB color space.

(3) Adjustment of DC Model radius

To improve the accuracy and adaptability of forest region extraction, the center and entire remote sensing picture acquired in the first phase is mapped or converted to HIS color space.Each pixel's Euclidean distance to the RSI center is calculated. The k-means clustering algorithm subdivide the forest into clusters and determines the Euclidean distance between the cluster center and the DC model's center, which is then used as the DC model's radius.

$$h = [h, s, i] \tag{20}$$

$$R = (h - h_0)^2 + (s - s_0)^2 + (i - i_0)^2 \tag{21}$$

$$\delta(i) = \sum_{k=1}^{k} |U_k^i - U_k^{i-1}| \tag{22}$$

P denotes the center of the DC model and the value would have been obtained by the histogram model in the RGB to HIS color scheme. R is the Euclidean distance and $\delta(i)$ represents

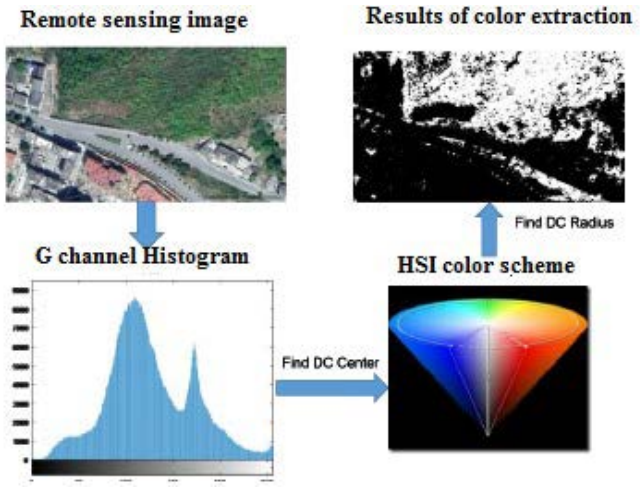


FIGURE 39. DC model in color extraction feature [126].

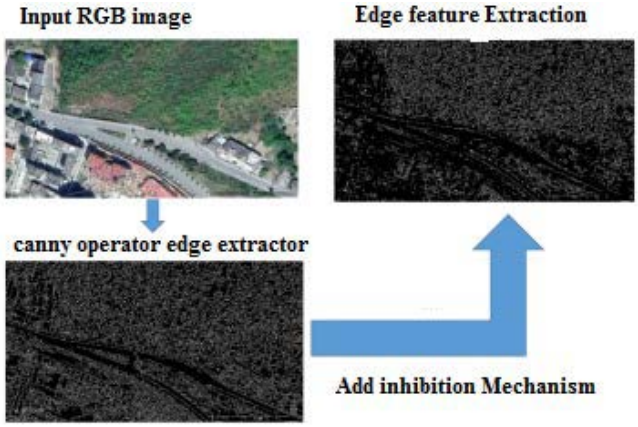


FIGURE 40. Edge feature extractor.

an is the iterations of the class algorithm. Figure 39 shows the color extraction feature of the DC model

(B) Edge Feature extraction

The goal of this procedure is to successfully eliminate non-forest areas. [126] proposed the canny operator as the edge detection operator because of its better performance than other operators in terms of edge feature detection. In particular, denoising is key for image processing, and in this particular instance, a Gaussian filter was employed to smoothen the image, thereby preserving the edges. The amplitude and direction of the gradient are then calculated using the finite difference of the step-wise derivative. The canny edge detector operator returns only the maximum value and uses the non-maximum suppression operation to suppress the field's conspicuous points, resulting in a corner point with high precision and clear vision. Finally, by using a dual threshold setting, discrete edges are linked together to form a continuous edge. Figure 35 shows the stages of an edge feature extractor.

IX. PERFORMANCE EVALUATION MATRIX

The major matrices to measure the performance of the model in forest image classification are: False Positive Rate (FPR), Accuracy (Acc), F1 score, Precision-Recall Curve, and

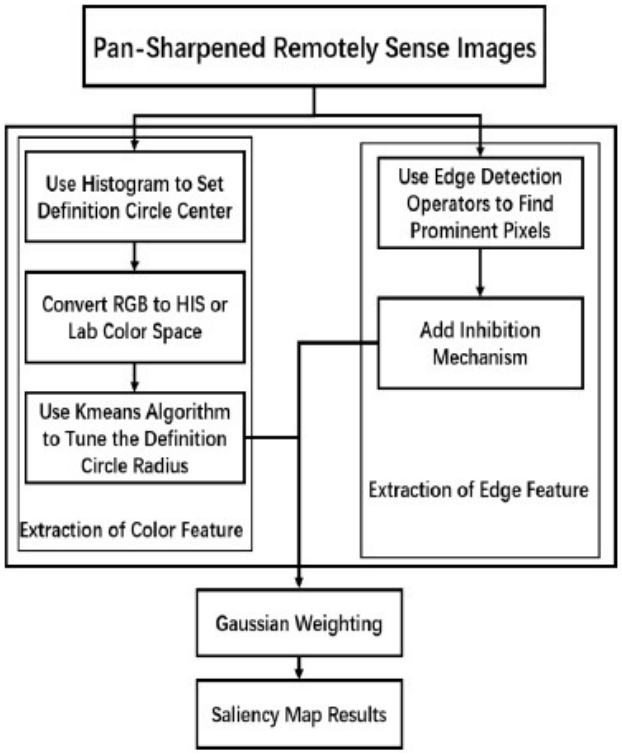


FIGURE 41. Flow diagram of the algorithm [126].

Average Precision (AP).

$$F1 = 2\frac{Precision * Recall}{Precision + Recall} \tag{23}$$

$$FRP = \frac{Numberofmisclassifiedforestimages}{Numberofimages} * 100\% \tag{24}$$

$$Acc = \frac{Numberofcorrectlyclassifiedimages}{Numberofimages} * 100\% \tag{25}$$

$$AP = areaunderthe\text{precision-recall}curvePRC \tag{26}$$

Measurements for image segmentation area evaluation are presented in Table 2. The Area Fitness Index (AFI) was proposed by [63] and the remaining measurements by [2]

The average distance between the reference object and its matching image object is described by the Position Discrepancy Index (PDI). The Overall PDI is the average of the PDI.

$$let\ a = (x(k) - X_r)^2 + (Y(k) - Y_r)^2 \tag{27}$$

$$let\ b = \sum_{i=1}^{M} \sqrt{(X(I) - X_r)^2} + (Y(I) - Y_r)^2) \tag{28}$$

$$PDI = \frac{1}{N+M}(\sum_{k=1}^{N} \sqrt{(a+b)} \tag{29}$$

$$PDI_{Overall} = \frac{1}{n}\sum_{i=n}^{n} PDI(i) \tag{30}$$

X. PERFORMANCE ANALYSIS OF THE STATE OF THE ART

Results based on the CNN with hyperparameter settings of patch size L = 15, regulation strength $\alpha = 0.001$, and

2. PREVIOUS WORKS: SURVEY PAPERS

TABLE 2. Evaluation Matrix.

Evaluation Matrix		
Measurement	Definition	Description
Area Fitness Index(AFI)	$\frac{A_r - A_{LargestImageObject}}{A_r}$	When AFI$>$ 0,over segmentation, When AFI $<$ 0,under segmentation
$OE_{Overall}$	$\frac{\sum_{i=1}^{n}(OE(i)*A_r(i)}{\sum_{i=1}^{n}A_r(i)}$	The weighted average of OE
Omission error (OE)	$\frac{\sum_{k=1}^{n_{exp}}A_e(k)-(A_e(k)\in A_r)}{A_r}$	Describes the under segmentation. An OE closer to zero means less under segmentation.
$CE_{Overall}$	$\frac{\sum_{i=1}^{n}(CE(i)*A_r(i))}{\sum_{i=1}^{n}A_r(i)}$	The weighted average of OE
Overall Area Discrepancy Index $ADI_{Overall}$	$\sqrt{OE^2_{Overall}+CE^2_{Overall}}$	The overall of area discrepancy index of over and under segmentation. When ADI is zero, segmentation quality is optimal

TABLE 3. Segmentation results based on PDI and ADI.

Species	Species Code	Hyperspectral			RGB		
		Precision	Recall	F-score	Precision	Recall	F-Score
White fir	0	0.76	0.81	0.78	0.46	0.53	0.49
Red fir	1	0.76	0.72	0.74	0.41	0.30	0.35
Incense cedar	2	0.90	0.85	0.88	0.50	0.44	0.47
Jeffrey pine	3	0.93	0.96	0.95	0.65	0.73	0.69
Sugar pine	4	0.90	0.96	0.93	0.67	0.68	0.67
Black oak	5	0.73	0.61	0.67	0.69	0.61	0.65
Lodgepole pine	6	0.84	0.87	0.86	0.54	0.47	0.50
Dead	7	0.90	0.85	0.88	0.88	0.86	0.87
Ave/Total		0.87	0.87	0.87	0.64	0.64	0.64

TABLE 4. Edge feature extractor.

Shape	Compactness	Scale	$OE_{overall}$ (%)	$CE_{overall}$ (%)	$ADI_{overall}$ (%)	$PDI_{overall}$ (m)
0.1	0.1	60	6.34	5.25	8.22	5.06
0.1	0.1	70	7.75	4.99	9.22	4.83
0.1	0.1	80	8.09	5.47	9.77	4.73
0.1	0.1	90	8.99	6.83	11.30	4.71
0.1	0.1	100	9.02	8.27	12.23	5.08
0.1	0.1	110	10.25	8.84	13.54	4.77
0.1	0.1	120	11.46	8.68	14.37	4.71
0.1	0.3	60	5.22	4.55	6.93	5.48
0.1	0.3	70	6.01	4.45	7.48	4.89
…	…	…	…	…	…	…
0.3	0.7	60	4.11	4.05	5.77	3.81
0.3	0.7	70	4.75	4.28	6.39	4.06
0.3	0.7	80	5.43	4.26	6.90	4.17
…	…	…	…	…	…	…
0.9	0.9	60	22.41	12.71	25.76	5.25
0.9	0.9	70	24.02	16.75	29.28	5.41
0.9	0.9	80	31.88	14.78	35.14	5.68
0.9	0.9	90	36.01	15.14	39.07	6.88
0.9	0.9	100	40.85	12.74	42.80	7.65
0.9	0.9	110	45.56	10.32	46.71	8.23
0.9	0.9	120	48.05	9.00	48.89	9.33

C = 32 filter kernels in the first convolutional layer up to a maximum of C' = 128 kernels. Using Tensorflow and Keras mechanisms, the final CNN classifiers used the hyperspectral imagery to outperform the RGB subset image as indicated by precision, recall, or F-score. Results are presented in Table 3.

Table 4 displays state-of-the-art segmentation results obtained using a supervised segmentation method and the following matrix measurements: AFI (index), OE, $OE_OVERALL$, CE, $CE_OVERALL$, ADI, PDI.

Object fate analysis and the method proposed by [63] do not objectively express segmentation quality results. Table 5 indicates that AFI ranges from 0.561 to −0.280 when shape and compactness are both at 0.1 and the scale parameter is changed from 60 to 120.

TABLE 5. Performance of CNN.

Metrics	Deviation percentage interval	5% sampling proportion		10% sampling proportion		15% sampling proportion		20% sampling proportion		25% sampling proportion		30% sampling proportion	
		^{1}T	^{2}CP (%)	^{1}T	^{2}CP (%)	^{1}T	^{2}CP (%	^{1}T	^{2}CP (%)	^{1}T	^{2}CP (%)	^{1}T	^{2}CP (%)
$ADI_{overall}$	<2%	16	32	23	46	29	58	37	74	39	78	46	92
	2–4%	11	54	13	72	9	76	13	100	11	100	4	100
	4–6%	9	72	10	92	11	98	0	100	0	100	0	100
	6–8%	6	84	2	96	1	100	0	100	0	100	0	100
	8–10%	5	94	2	100	0	100	0	100	0	100	0	100
	>10%	3	100	0	100	0	100	0	100	0	100	0	100
$PDI_{overall}$	<2%	13	26	17	34	27	54	33	66	41	82	44	88
	2–4%	14	54	14	62	8	70	17	100	9	100	6	100
	4–6%	5	64	6	74	9	88	0	100	0	100	0	100
	6–8%	9	82	8	90	3	94	0	100	0	100	0	100
	8–10%	8	98	5	100	3	100	0	100	0	100	0	100
	>10%	1	100	0	100	0	100	0	100	0	100	0	100

XI. RECOMMENDATION

Pixel-based techniques have been commonly used for image analysis and classification for a very long time. However, due to the massive growth of high spatial resolution images and the fact that pixel based methods only work with spectral information, the technique could not be fully utilized because it does not incorporate spatial, texture, and shape information, [127]. Previous studies have also shown that such approaches cause noise in the output message, otherwise known as the "salt and pepper effect." [128]. Due to the limitations of traditional pixel-based methods to cope with high-resolution imagery, OBIA methods have become increasingly popular because they have a high degree of information utilization, strong anti-interference, a high degree of data integration, and high classification accuracy [129], [130]. However, GEOBIA techniques are made up of knowledge and rules purely from domain expert knowledge, such that they enhance the subjectivity of image interpretation processes. Given the evolution of remote sensing science as a result of artificial intelligence, this study suggests that we pay more attention to Good Old-Fashioned Artificial Intelligence (GOFAI), which is based on sound mathematics and logic to construct symbolic representations of abstract notions [1]. This research highly recommends a shift towards remote sensing image analysis with ontologies because such technology allows management, aggregation, and sharing of the knowledge of remote sensing and domain experts. Formal ontologies explicitly define expert knowledge that is used to interpret remote sensing images. This improves the sharing and reuse of formalized remote sensing expert knowledge.

XII. CONCLUSION

This paper is a critical and analytical survey of the methods for forest image detection and classification. It is a comprehensive review of the techniques used to detect objects of interest in an image that will be analysed for classification of forests. These techniques cover semantic segmentation techniques, feature extraction methods and finally classification techniques. Exploration of knowledge based approaches in form of GEOBIA were analysed and how their shortcoming in terms of dual mode of defining geographic concept, vagueness and ambiguity of geographic concepts, and semantic gaps were addressed by ontology knowledge based approaches. Performance of the state of the art Tensorflow and Keras for image classification were analysed. Formal ontologies knowledge representation was recommended for state of the art approach for detecting objects of interest. CNN methods for semantic segmentation were critically analysed and these were; AlexNet, VGGNet, GoogLeNet, FCN, UNet, SegNet, DeepNet and ResNet.

REFERENCES

[1] D. Arvor, M. Belgiu, Z. Falomir, I. Mougenot, and L. Durieux, "Ontologies to interpret remote sensing images: Why do we need them?" *GIScience Remote Sens.*, vol. 56, no. 6, pp. 911–939, Aug. 2019.

[2] J. Cheng, Y. Bo, Y. Zhu, and X. Ji, "A novel method for assessing the segmentation quality of high-spatial resolution remote-sensing images," *Int. J. Remote Sens.*, vol. 35, no. 10, pp. 3816–3839, May 2014.

[3] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 177, pp. 11–28, Jul. 2016.

[4] R. M. Dufour, E. L. Miller, and N. P. Galatsanos, "Template matching based object recognition with unknown geometric parameters," *IEEE Trans. Image Process.*, vol. 11, no. 12, pp. 1385–1396, Dec. 2002.

[5] K. Bahareh, M. Shattri, S. Helmi, and H. Alfian, "Integration of template matching and object-based image analysis for semi-automatic oil palm tree counting in UAV images," in *Proc. 37th Asian Conf. Remote Sens. (ACRS)*, vol. 3, 2016, pp. 2333–2340.

[6] I. A. Aljarrah and A. S. Ghorab, "Object recognition system using template matching based on signature and principal component analysis," *Int. J. Digit. Inf. Wireless Commun.*, vol. 2, no. 2, pp. 156–163, 2012.

[7] I. Jordi, V. Aurthur, M. Arias, B. Tardy, D. Morin, and I. Rodes, "Operational high resolution land cover map production at the country scale using satellite image time series," *Remote Sens.*, vol. 9, no. 1, p. 95, 2017.

[8] M. Cristina, A. Picoli, G. Camara, I. Sanches, R. Simões, A. Carvalho, A. Maciel, A. Coutinho, J. Esquerdo, J. Antunes, R. Anzolin, D. Arvor, and C. Almeida, "Big earth observation time series analysis for monitoring Brazilian agriculture," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 328–339, Nov. 2018.

[9] M. Papadomanolaki, M. Vakalopoulou, S. Zagoruyko, and K. Karantzalos, "Benchmarking deep learning frameworks for the classification of very high resolution satellite multispectral data," *ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 7, pp. 83–88, Jun. 2016.

[10] J. Campbell, *Introduction to Remote Sensing*, 3rd ed. New York, NY, USA: Guilford Press, 2002.

[11] G. Marcus, "Deep learning: A critical appraisal," 2018, *1801.00631*.

[12] A. L. Ali, Z. Falomir, F. Schmid, and C. Freksa, "Rule-guided human classification of volunteered geographic information," *ISPRS J. Photogramm. Remote Sens.*, vol. 127, pp. 3–15, May 2017, doi: 10.1016/j.isprsjprs.2016.06.003.

[13] T. R. Martha, N. Kerle, C. J. V. Westen, V. Jetten, and K. V. Kumar, "Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4928–4943, Dec. 2011.

[14] D. Chaudhuri, N. K. Kushwaha, and A. Samal, "Semi-automated road detection from high resolution satellite images by directional morphological enhancement and segmentation techniques," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1538–1544, Oct. 2012.

[15] A. H. S. Solberg, "Contextual data fusion applied to forest map revision," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1234–1243, May 1999.

[16] A. Hanif, A. B. Mansoor, and A. S. Imran, *Performance Analysis of Vehicle Detection Techniques: A Concise Survey*, vol. 746. Cham, Switzerland: Springer, 2018, doi: 10.1007/978-3-319-77712-2_46.

[17] G. Chen, X. Zhang, Q. Wang, F. Dai, Y. Gong, and K. Zhu, "Symmetrical dense-shortcut deep fully convolutional networks for semantic segmentation of very-high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1633–1644, May 2018.

[18] R. Mohan and R. Nevatia, "Using perceptual organization to extract 3D structures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 11, pp. 1121–1139, Nov. 1989.

[19] H. G. Akcay and S. Aksoy, "Building detection using directional spatial constraints," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2010, pp. 1932–1935.

[20] J. Peng and Y. Liu, "Model and context-driven building extraction in dense urban aerial images," *Int. J. Remote Sens.*, vol. 26, no. 7, pp. 1289–1307, 2005.

[21] A. O. Ok, C. Senaras, and B. Yuksel, "Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1701–1717, Mar. 2013.

[22] Y.-T. Liow and T. Pavlidis, "Use of shadows for extracting buildings in aerial images," *Comput. Vis., Graph., Image Process.*, vol. 49, no. 2, pp. 242–277, Feb. 1990.

[23] X. Zhang, X. Feng, P. Xiao, G. He, and L. Zhu, "Segmentation quality evaluation using region-based precision and recall measures for remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 102, pp. 73–84, Apr. 2015.

[24] T. Su and S. Zhang, "Object-based crop classification in hetao plain using random forest," *Earth Sci. Informat.*, vol. 14, no. 1, pp. 119–131, Mar. 2021.

[25] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 929–944, Jun. 2007.

[26] Y. Wang, Q. Qi, and Y. Liu, "Unsupervised segmentation evaluation using area-weighted variance and jeffries-matusita distance for remote sensing images," *Remote Sens.*, vol. 10, no. 8, p. 1193, Jul. 2018.

[27] H. Gao, Y. Tang, L. Jing, H. Li, and H. Ding, "A novel unsupervised segmentation quality evaluation method for remote sensing images," *Sensors*, vol. 17, no. 10, p. 2427, Oct. 2017.

[28] G. Modica, G. De Luca, G. Messina, and S. Praticò, "Comparison and assessment of different object-based classifications using machine learning algorithms and UAVs multispectral imagery: A case study in a citrus orchard and an onion crop," *Eur. J. Remote Sens.*, vol. 54, no. 1, pp. 431–460, Jan. 2021.

[29] G. J. Hay and G. Castilla, "Geographic object-based image analysis (GEOBIA): A new name for a new discipline," in *Object-Based Image Analysis*. Berlin, Germany: Springer, 2008, pp. 75–89.

[30] J. Krishnaswamy, M. C. Kiran, and K. N. Ganeshaiah, "Tree model based eco-climatic vegetation classification and fuzzy mapping in diverse tropical deciduous ecosystems using multi-season NDVI," *Int. J. Remote Sens.*, vol. 25, no. 6, pp. 1185–1205, Mar. 2004.

[31] K. J. Feeley, T. W. Gillespie, and J. W. Terborgh, "The utility of spectral indices from Landsat ETM+ for measuring the structure and composition of tropical dry forests," *Biotropica: J. Biol. Conservation*, vol. 37, no. 4, pp. 508–519, 2005.

[32] G. A. Sanchez-Azofeifa, K. L. Castro, B. Rivard, M. R. Kalascka, and R. C. Harriss, "Remote sensing research priorities in tropical dry forest environments," *Biotropica*, vol. 35, no. 2, pp. 134–142, Jun. 2003.

[33] S. Martinuzzi, W. A. Gould, O. M. Ramos González, A. Martínez Robles, P. Calle Maldonado, N. Pérez Buitrago, and J. J. Fumero Cabán, "Mapping tropical dry forest habitats integrating landsat NDVI, ikonos imagery, and topographic information in the Caribbean island of Mona," *Revista de Biología Tropical*, pp. 625–639, Nov. 2006.

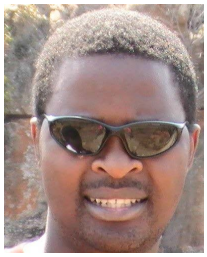
[34] B. Bennett, "What is a forest? On the vagueness of certain geographic concepts," in *Proc. TOPOI*, 2002, pp. 2–17.

[35] B. Bennett, "Foundations for an ontology of environment and habitat," in *Proc. FOIS*, 2010, pp. 31–44.
[36] A. Mayamba, R. M. Byamungu, B. V. Broecke, H. Leirs, P. Hieronimo, A. Nakiyemba, M. Isabirye, D. Kifumba, D. N. Kimaro, M. E. Mdangi, and L. S. Mulungu, "Factors influencing the distribution and abundance of small rodent pest species in agricultural landscapes in eastern Uganda," *J. Vertebrate Biol.*, vol. 69, no. 2, Oct. 2020, Art. no. 020002.
[37] H. G. Lund, "When is a forest not a forest?" *J. Forestry*, vol. 100, no. 8, pp. 21–28, 2002.
[38] E. Romijn, J. H. Ainembabazi, A. Wijaya, M. Herold, A. Angelsen, L. Verchot, and D. Murdiyarso, "Exploring different forest definitions and their impact on developing REDD+ reference emission levels: A case study for Indonesia," *Environ. Sci. Policy*, vol. 33, pp. 246–259, Nov. 2013.
[39] C. E. Woodcock and A. H. Strahler, "The factor of scale in remote sensing," *Remote Sens. Environ.*, vol. 21, no. 3, pp. 311–332, Apr. 1987.
[40] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.
[41] C. Unger and P. Cimiano, "Pythia: Compositional meaning construction for ontology-based question answering on the semantic web," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Lecture Notes in Computer Science), vol. 6716. Germany: Bielefeld Univ., 2011, pp. 153–160.
[42] R. H. Kilmann and K. W. Thomas, "Developing a forced-choice measure of conflict-handling behavior: The 'MODE' instrument," *Educ. Psychol. Meas.*, vol. 37, no. 2, pp. 309–325, 1977.
[43] B. Bachimont, A. Isaac, and R. Troncy, "Semantic commitment for designing ontologies: A proposal," in *Proc. Int. Conf. Knowl. Eng. Knowl. Manage.* Berlin, Germany: Springer, 2002, pp. 114–121.
[44] E. F. Fama and M. C. Jensen, "Separation of ownership and control," *J. Law Econ.*, vol. 26, no. 2, pp. 301–325, 1983.
[45] K. Satoh, *Nonmonotonic Reasoning by Minimal Belief Revision*. Tokyo, Japan: ICOT Research Center (Institute for New Generation Computer Technology), 1988.
[46] T. Gruber, "What is an ontology," Stanford Univ., Stanford, CA, USA, Tech. Rep. KSL92-71, 1993.
[47] S. Andrés, D. Arvor, I. Mougenot, T. Libourel, and L. Durieux, "Ontology-based classification of remote sensing images using spectral rules," *Comput. Geosci.*, vol. 102, pp. 158–166, May 2017.
[48] D. Mallenby, "Handling vagueness in ontologies of geographical information," Ph.D. dissertation, School Comput., Univ. Leeds, Leeds, U.K., 2008. [Online]. Available: http://etheses.whiterose.ac.U.K./1373/
[49] N. Eric Maillot and M. Thonnat, "Ontology based complex object recognition," *Image Vis. Comput.*, vol. 26, no. 1, pp. 102–113, Jan. 2008.
[50] C. Eschenbach and M. Grüninger, "Formal ontology in information systems," in *Proc. 5th Int. Conf. (FOIS)*, vol. 110, 2008, pp. 68–71.
[51] M. Davis, S. King, N. Good, and R. Sarvas, "From context to content: Leveraging context to infer media metadata," in *Proc. 12th Annu. ACM Int. Conf. Multimedia*, 2004, pp. 188–195.
[52] F. Nack, C. Dorai, and S. Venkatesh, "Computational media aesthetics: Finding meaning beautiful," *IEEE MultimediaMag.*, vol. 8, no. 4, pp. 10–12, Oct. 2001.
[53] H. Gu, H. Li, L. Yan, Z. Liu, T. Blaschke, and U. Soergel, "An object-based semantic classification method for high resolution remote sensing imagery using ontology," *Remote Sens.*, vol. 9, no. 4, p. 329, 2017.
[54] A. Baraldi, V. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino, "Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+ images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2563–2586, Sep. 2006.
[55] A. M. Arifjanov, S. B. Akmalov, T. U. Apakhodjaeva, and D. S. Tojikhodjaeva, "Comparison of pixel to pixel and object based image analysis using worldview-2 satellite images of vangiobod village of syndria province," *Remote Methods Earth Res.* vol. 26, no. 2, pp. 313–321, 2020.
[56] N. Durand, S. Derivaux, G. Forestier, C. Wemmert, P. Gançarski, O. Boussaid, and A. Puissant, "Ontology-based object recognition for remote sensing image interpretation," in *Proc. 19th IEEE Int. Conf. Tools Artif. Intell. (ICTAI)*, vol. 1, Oct. 2007, pp. 472–479.
[57] S. R. Phinn, C. M. Roelfsema, and P. J. Mumby, "Multi-scale, object-based image analysis for mapping geomorphic and ecological zones on coral reefs," *Int. J. Remote Sens.*, vol. 33, no. 12, pp. 3768–3797, Jun. 2012.
[58] B. Gajderowicz, "Using decision trees for inductively driven semantic integration and ontology matching," M.S. thesis, Dept. Comput. Sci., Ryerson Univ., Toronto, ON, Canada, 2011.
[59] B. Gajderowicz and A. Sadeghian, "Ontology granulation through inductive decision trees," in *Proc. URSW*, 2009, pp. 39–50.
[60] N. Kartha and A. Novstrup, "Ontology and rule based knowledge representation for situation management and decision support," *Proc. SPIE*, vol. 7352, May 2009, Art. no. 73520P.
[61] J. C. Giarratano and G. D. Riley, *Expert Systems: Principles and Programming*. Pacific Grove, CA, USA: Brooks/Cole, 2005.
[62] D. A. Waterman, D. B. Lenat, and F. Hayes-Roth, *Building Expert Systems*. Reading, MA, USA: Addison-Wesley, 1983.
[63] P. He, "Counter cyber attacks by semantic networks," in *Emerging Trends in ICT Security*. Amsterdam, The Netherlands: Elsevier, 2014, pp. 455–467.
[64] G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, and R. Rosati, "Using ontologies for semantic data integration," in *A Comprehensive Guide Through Italian Database Res. Over Last 25 Years*. Cham, Switzerland: Springer, 2018, pp. 187–202.
[65] F. Baader, *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[66] B. C. Grau, I. Horrocks, B. Motik, B. Parsia, P. Patel-Schneider, and U. Sattler, "OWL 2: The next step for OWL," *J. Web Semantics*, vol. 6, no. 4, pp. 309–322, Nov. 2008.
[67] B. C. Grau, I. Horrocks, Y. Kazakov, and U. Sattler, "A logical framework for modularity of ontologies," in *Proc. IJCAI*, 2007, pp. 298–303.
[68] S. Ghilardi, C. Lutz, and F. Wolter, "Did I damage my ontology," in *Proc. KR*, 2006, pp. 187–197.
[69] S. Roy and I. J. Cox, "A maximum-flow formulation of the N-camera stereo correspondence problem," in *Proc. 6th Int. Conf. Comput. Vis.*, Jan. 1998, pp. 492–499.
[70] R. Geerken, B. Zaitchik, and J. P. Evans, "Classifying rangeland vegetation type and coverage from NDVI time series using Fourier filtered cycle similarity," *Int. J. Remote Sens.*, vol. 26, no. 24, pp. 5535–5554, Dec. 2005.
[71] Y. Guo, S. Han, Y. Li, C. Zhang, and Y. Bai, "K-nearest neighbor combined with guided filter for hyperspectral image classification," *Proc. Comput. Sci.*, vol. 129, pp. 159–165, Jan. 2018.
[72] G. De Luca, J. M. N. Silva, S. Cerasoli, J. Araújo, J. Campos, S. Di Fazio, and G. Modica, "Object-based land cover classification of cork oak woodlands using UAV imagery and orfeo toolbox," *Remote Sens.*, vol. 11, no. 10, p. 1238, May 2019.
[73] A. D. P. Pacheco, J. A. D. S. Junior, A. M. Ruiz-Armenteros, and R. F. F. Henriques, "Assessment of K-nearest neighbor and random forest classifiers for mapping forest fire areas in central Portugal using Landsat-8, Sentinel-2, and Terra imagery," *Remote Sens.*, vol. 13, no. 7, p. 1345, Apr. 2021.
[74] P. T. Noi and M. Kappas, "Comparison of random forest, K-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery," *Sensors*, vol. 18, no. 1, p. 18, 2018.
[75] E. Tomppo, M. Haakana, M. Katila, and J. Peräsaari, *Multi-Source National Forest Inventory: Methods and Applications*, vol. 18. Springer, 2008.
[76] L. Tlig, M. Bouchouicha, M. Tlig, M. Sayadi, and E. Moreau, "A fast segmentation method for fire forest images based on multiscale transform and PCA," *Sensors*, vol. 20, no. 22, p. 6429, Nov. 2020.
[77] S. M. De Jong and F. D. Van der Meer, *Remote Sensing Image Analysis: Including the Spatial Domain*, vol. 5. Springer, 2007.
[78] D. Kaur and Y. Kaur, "Various image segmentation techniques: A review," *Int. J. Comput. Sci. Mobile Comput.*, vol. 3, no. 5, pp. 809–814, 2014.
[79] Y.-J. Zhang, "An overview of image and video segmentation in the last 40 years," in *Advances in Image and Video Segmentation*. Dordrecht, The Netherlands: 2006, pp. 1–16.
[80] T. Lindeberg and M.-X. Li, "Segmentation and classification of edges using minimum description length approximation and complementary junction cues," *Comput. Vis. Image Understand.*, vol. 67, no. 1, pp. 88–98, Jul. 1997.
[81] S. Yuheng and Y. Hao, "Image segmentation algorithms overview," 2017, *arXiv:1707.02051*.
[82] N. Senthilkumaran and R. Rajesh, "Image segmentation—A survey of soft computing approaches," in *Proc. Int. Conf. Adv. Recent Technol. Commun. Comput.* Stockholm, Sweden: KTH (Roy. Inst. Technol.), Oct. 2009, pp. 844–846.
[83] M. K. Kundu and S. K. Pal, "Thresholding for edge detection using human psychovisual phenomena," *Pattern Recognit. Lett.*, vol. 4, no. 6, pp. 433–441, 1986.

[84] M. R. Khokher, A. Ghafoor, and A. M. Siddiqui, "Image segmentation using multilevel graph cuts and graph development using fuzzy rule-based system," *IET Image Process.*, vol. 7, no. 3, pp. 201–211, 2013.
[85] T. Blaschke, C. Burnett, and A. Pekkarinen, "Image segmentation methods for object-based analysis and classification," in *Remote Sensing Image Analysis: Including The Spatial Domain*. Dordrecht, The Netherlands: Springer, 2004, pp. 211–236.
[86] T. Lei, X. Jia, Y. Zhang, S. Liu, H. Meng, and A. K. Nandi, "Superpixel-based fast fuzzy C-means clustering for color image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 9, pp. 1753–1766, Sep. 2019.
[87] P. Neubert and P. Protzel, "Compact watershed and preemptive SLIC: On improving trade-offs of superpixel segmentation algorithms," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 996–1001.
[88] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," *Expert Syst. Appl.*, vol. 169, May 2021, Art. no. 114417.
[89] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1097–1105.
[90] B. Liu, X. Yu, A. Yu, and G. Wan, "Deep convolutional recurrent neural network with transfer learning for hyperspectral image classification," *J. Appl. Remote Sens.*, vol. 12, no. 2, 2018, Art. no. 026028.
[91] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2414–2423.
[92] O. A. B. Penatti, K. Nogueira, and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 44–51.
[93] M. Volpi and V. Ferrari, "Semantic segmentation of urban scenes by learning local class interactions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 1–9.
[94] M. Lin and Q. Yan, "Network in network," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 1–4.
[95] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, 2016.
[96] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2015, pp. 234–241.
[97] J. E. Ball, D. T. Anderson, and C. S. Chan, "Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community," *J. Appl. Remote Sens.*, vol. 11, no. 4, 2017, Art. no. 042609.
[98] S. Chilamkurthy. (2017). *A 2017 Guide to Semantic Segmentation with Deep Learning*. [Online]. Available: https://blog.qure.ai/notes/semantic-segmentation-deep-learning-review
[99] J. Le. (2017). *How to do Semantic Segmentation Using Deep Learning*. [Online]. Available: https://nanonets.com/blog/how-to-do-semantic-segmentation-using-deep-learning/
[100] A. Mittal. (2019). *Introduction to U-Net and Res-Net for Image Segmentation*. [Online]. Available: https://aditi-mittal.medium.com/introduction-to-u-net-and-res-net-for-image-segmentation-9afcb432ee2f
[101] S. Kentsch, M. L. Lopez Caceres, D. Serrano, F. Roure, and Y. Diez, "Computer vision and deep learning techniques for the analysis of drone-acquired forest images, a transfer learning study," *Remote Sens.*, vol. 12, no. 8, p. 1287, Apr. 2020.
[102] M. Šulc, D. Mishkin, and J. Matas, "Very deep residual networks with maxout for plant identification in the wild," in *Proc. Work. Notes CLEF*, 2016, pp. 1–8.
[103] M. Onishi and T. Ise, "Automatic classification of trees using a UAV onboard camera and deep learning," 2018, *arXiv:1804.10390*.
[104] S. Natesan, C. Armenakis, and U. Vepakomma, "ResNet-based tree species classification using UAV images," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 475–481, Jun. 2019.
[105] M. Dyrmann, H. Karstoft, and H. S. Midtiby, "Plant species classification using deep convolutional neural network," *Biosyst. Eng.*, vol. 151, pp. 72–80, Nov. 2016.
[106] M. Onishi and T. Ise, "Automatic classification of trees using a UAV onboard camera and deep learning," 2018, *arXiv:1804.10390*.
[107] O. Brovkina, E. Cienciala, P. Surový, and P. Janata, "Unmanned aerial vehicles (UAV) for assessment of qualitative classification of Norway spruce in temperate forest stands," *Geo-spatial Inf. Sci.*, vol. 21, no. 1, pp. 12–20, Jan. 2018.
[108] M. J. Zimmer-Gembeck and M. Helfand, "Ten years of longitudinal research on U.S. adolescent sexual behavior: Developmental correlates of sexual intercourse, and the importance of age, gender and ethnic background," *Developmental Rev.*, vol. 28, no. 2, pp. 153–224, 2008.
[109] S. Watanabe, K. Sumi, and T. Ise, "Identifying the vegetation type in Google Earth images using a convolutional neural network: A case study for Japanese bamboo forests," *BMC Ecol.*, vol. 20, no. 1, pp. 1–14, Dec. 2020.
[110] E. Guirado, S. Tabik, D. Alcaraz-Segura, J. Cabello, and F. Herrera, "Deep-learning versus OBIA for scattered shrub detection with Google Earth imagery: Ziziphus lotus as case study," *Remote Sens.*, vol. 9, no. 12, p. 1220, Nov. 2017.
[111] P. P. Ippolito. (2019). *Feature Extraction Techniques*. Accessed: Apr. 29, 2020. [Online]. Available: https://towardsdatascience.com/feature-extraction-techniques-d619b56e31be
[112] A. Ghodsi, "Dimensionality reduction a short tutorial," Ph.D. dissertation, Dept. Statist. Actuarial Sci., Univ. Waterloo, Waterloo, ON, Canada, 2006, vol. 37, no. 38.
[113] B. Ghojogh, M. N. Samad, S. Asif Mashhadi, T. Kapoor, W. Ali, F. Karray, and M. Crowley, "Feature selection and feature extraction in pattern analysis: A literature review," 2019, *arXiv:1905.02845*.
[114] C. Citro, "rules. In Proc. 20th Int. Conf. very large data bases, VLDB, volume 1215, pages 487–499, 1994.[5] Alfred V. Aho, Ravi Sethi and Jeffrey D. Ullman. Compilers: Principles, techniques, and tools. Boston, MA: Addison-Wesley, 1986.[6] Adrian Akmajian, Ann K. Farmer, Lee Bickmore, Richard A. Demers and," *Learning*, vol. 5, no. 1, pp. 71–99, 1990.
[115] C. A. Brooks and K. Iagnemma, "Vibration-based Terrain classification for planetary exploration rovers," *IEEE Trans. Robot.*, vol. 21, no. 6, pp. 1185–1191, Dec. 2005.
[116] F. Subhan, S. Saleem, H. Bari, W. Z. Khan, S. Hakak, S. Ahmad, and A. M. El-Sherbeeny, "Linear discriminant analysis-based dynamic indoor localization using Bluetooth low energy (BLE)," *Sustainability*, vol. 12, no. 24, p. 10627, Dec. 2020.
[117] Y. Mo, Z. Zhang, Y. Lu, W. Meng, and G. Agha, "Random forest based coarse locating and KPCA feature extraction for indoor positioning system," *Math. Problems Eng.*, vol. 2014, Oct. 2014, Art. no. 850926.
[118] L. C. Marsh and D. R. Cormier, *Spline Regression Models*, no. 137. Newbury Park, CA, USA: Sage, 2001.
[119] L. K. Saul and S. T. Roweis, "Think globally, fit locally: Unsupervised learning of low dimensional manifolds," *J. Machine Learn. Res.*, vol. 4, pp. 119–155, Jun. 2003.
[120] M. Li, X. Luo, J. Yang, and Y. Sun, "Applying a locally linear embedding algorithm for feature extraction and visualization of MI-EEG," *J. Sensors*, vol. 2016, Aug. 2016, Art. no. 7481946.
[121] G. Hinton and S. T. Roweis, "Stochastic neighbor embedding," in *Proc. NIPS*, vol. 15, 2002, pp. 833–840.
[122] S. Kullback, *Information Theory and Statistics*. Chelmsford, MA, USA: Courier Corporation, 1997.
[123] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne," *J. Mach. Learn. Res.*, vol. 9, no. 11, 2008.
[124] D. Gu, Z. Han, and Q. Wu, "Feature extraction to polar image," *J. Comput. Commun.*, vol. 5, no. 11, pp. 16–26, 2017.
[125] F. Alamdar and M. Keyvanpour, "A new color feature extraction method based on QuadHistogram," *Proc. Environ. Sci.*, vol. 10, pp. 777–783, Jan. 2011.
[126] H. Du and Y. Zhuang, "Optical remote sensing images feature extraction of forest regions," in *Proc. IEEE Int. Conf. Signal, Inf. Data Process. (ICSIDP)*, Dec. 2019, pp. 1–5.
[127] H. Luo, L. Li, H. Zhu, X. Kuai, Z. Zhang, and Y. Liu, "Land cover extraction from high resolution ZY-3 satellite imagery using ontology-based method," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 3, p. 31, 2016.
[128] O. Oke Alice, O. Omidiora Elijah, A. Fakolujo Olaosebikan, S. Falohun Adeleye, and S. Olabiyisi, "Effect of modified Wiener algorithm on noise models," *Int. J. Eng. Technol.*, vol. 2, no. 8, pp. 1024–1033, 2012.
[129] G. Hay and G. Castilla, "Object-based image analysis: Strengths, weaknesses, opportunities and threats (SWOT)," in *Proc. 1st Int. Conf. (OBIA)*, 2006, pp. 4–5.
[130] D. C. Duro, S. E. Franklin, and M. G. Dubé, "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery," *Remote Sens. Environ.*, vol. 118, pp. 259–272, Mar. 2012.

2. PREVIOUS WORKS: SURVEY PAPERS

CLOPAS KWENDA received the B.Sc. degree (Hons.) in computer science from the Bindura University of Science Education (BUSE), Zimbabwe, and the M.Sc. degree in computer science from the University of Zimbabwe (UZ), Zimbabwe. He is currently pursuing the Ph.D. degree with the University of KwaZulu-Natal (UKZN), South Africa. His research interests include image processing, artificial intelligence, machine learning, deep learning, and ontology building.



MANDLENKOSI GWETU received the Ph.D. degree in computer science (CS), specializing in medical image processing, from University of KwaZulu-Natal (UKZN), South Africa. He is a Senior Lecturer with UKZN. He is currently the Academic Leader of CS with UKZN. He is the Principal Investigator of the UKZN node in the Erasmus+ funded the Living Laboratories for Climate Change Multi-National Project and is an Alumni of the Heidelberg Laureate Forum. His research interests include deep learning, pattern recognition, and computer vision.

JEAN VINCENT FONOU DOMBEU received the B.Sc. degree (Hons.) in computer science from the University of Yaoundé I, Cameroonis, the M.Sc. degree in computer science from the University of KwaZulu-Natal, South Africa, and the Ph.D. degree in computer science from the North-West University, South Africa. He is a Senior Lecturer with the Department of Computer Science, University of KwaZulu-Natal (UKZN). His research interests include ontology engineering, semantic web, and machine learning—specifically, in ontology building, learning, modularization, ranking, summarization and visualization, artificial intelligence, machine learning and data mining methods for the semantic web, knowledge representation and reasoning on the web, and knowledge graphs and deep semantics.

2.1.2 Conclusion

The chapter pre-empted the challenges faced by object detection algorithms and recommended the integration of ontologies as a way of alleviating challenges faced by modern detection algorithms. The paper presented the state-of-the-art CNN-based model for satellite forest image classification.

2.2 A Critical Survey of GEOBIA Methods for forest Image Detection and Classification

2.2.1 Introduction

This section is an extension of the paper presented in section 2.1. This paper will critically examine previous studies that used the GEOBIA technique to process satellite forest images. Also, opportunities for Improving GEOBIA methods will be looked into. This review will serve as the basis for guidance in choosing the appropriate technique to settle for when it comes to the processing of satellite forest images. Recent methods that adopted ontologies for satellite forest image classification will also be critically examined.

This paper has been published by GEOCARTO INTERNATION journal (Taylor and Francis Journal).

Geocarto International

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgei20

A critical survey of GEOBIA methods for forest image detection and classification

Clopas Kwenda, Mandlenkosi Victor Gwetu & Jean Vincent Fonou-Dombeu

To cite this article: Clopas Kwenda, Mandlenkosi Victor Gwetu & Jean Vincent Fonou-Dombeu (2023) A critical survey of GEOBIA methods for forest image detection and classification, Geocarto International, 38:1, 2256302, DOI: 10.1080/10106049.2023.2256302

To link to this article: https://doi.org/10.1080/10106049.2023.2256302

GEOCARTO INTERNATIONAL
2023, VOL. 38, NO. 1, 2256302
https://doi.org/10.1080/10106049.2023.2256302

Taylor & Francis
Taylor & Francis Group

A critical survey of GEOBIA methods for forest image detection and classification

Clopas Kwenda, Mandlenkosi Victor Gwetu and Jean Vincent Fonou-Dombeu

School of Computer Science, Statistics and Mathematics, University of KwaZulu-Natal, Pietermaritzburg, South Africa

ABSTRACT

Modern earth observation sensors have revolutionized the remote sensing community by improving remote sensing image quality. However, Pixel-based image analysis methods have challenges in handling very high-resolution (VHR) imagery. Geographic Based Image Analysis (GEOBIA) yielded promising results, but it is not inflexible in capturing domain experts' expressions, therefore geographic information system professionals shifted to ontologies for remote sensing science. This paper advocates for the adoption of knowledge representation using ontologies in remote sensing. To this end, a survey of GEOBIA studies for image analysis and classification is presented, and the limitations of existing methods in reaching the remote sensing expert-level expectation are clarified. New GEOBIA development techniques as well as opportunities for improving GEOBIA models have been looked into. Recent studies that adopted ontologies in forest image classification are analyzed and recommendations for the remote sensing science community are provided, to highlight the advantages of ontologies in interpreting satellite images.

ARTICLE HISTORY
Received 13 June 2023
Accepted 2 September 2023

KEYWORDS
GEOBIA; machine learning; segmentation; remote sensing; ontology

1. Introduction

The launch of the first civilian satellite for earth observation (Landsat-1) has significantly transformed the remote sensing science community (Castilla and Hay 2008) by making it possible to acquire near real-time high-quality satellite imaging on demand. Remote sensing substantially simplifies the automated study of urban, suburban, and natural environments for applications such as monitoring urban expansion, detecting changes, crop prediction, forestation/deforestation, surveillance, human activities, mining, and so on Qin and Liu (2022). Satellite earth observation sensors coupled with the evolution of web services have tremendously improved access to satellite images, and principal agencies such as the National Aeronautics and Space Administration (NASA), United States Geological Survey (USGS), Brazilian National Institute for Space Research (INPE), Group on Earth Observation (GEO) and so forth, have ensured that large amounts of data are

CONTACT Clopas Kwenda 221072651@stu.ukzn.ac.za

freely available to the users (Arvor et al. 2013). The advent of modern remote sensors has improved the quality of remote sensing images which are made available to the community of users. Supervised pixel-based methods have been widely used for tasks relating to change detection in land use as well as land cover multi-temporal mapping Lu et al. (2013). However, the traditional pixel-based methods are unable to handle images produced by very high resolution (VHR) satellite imagery sensors (Castilla and Hay 2008), that is, (VHR <5 m) pixel size. Pixel-based methods have been hugely criticized for putting focus on presenting information as a digital number, i.e. how bright each pixel in an image is and it does not have the power to give details relating to spatial concepts of neighborhood, homogeneity, and proximity Souza-Filho et al. (2018). Machine learning (ML) is a sub-branch of artificial intelligence, and algorithms under ML are designed in such a way that they will be able to learn from data in order to predict corresponding outputs. Land cover classification has gained popular research in remote sensing, where both pixel-level classification and boundary mapping are all considered. Machine learning classifiers such as classification and regression trees (CART) Xiang et al. (2008), Random Forest (RF) Breiman (2001), and Support Vector Machines (SVM) Cortes and Vapnik (1995), have proved to perform better, and therefore have been widely used in land cover classification. The CART works by predicting a target variable using decision rules inferred from data features. The advantages of CART for land cover applications include its simple, explicit, and intuitive classification structure based on a set of 'if-then' rules. It can also be trained with any set of inputs without the need to adjust any parameters because by nature CART is a non-parametric model. CART was the first machine classifier to be used in land cover classification Pal and Mather (2003). However, CART has got issues with regard to computational complexity, high variable correlation, and noises from data collection and calibration, which have the potential to negatively affect classification and efficiency Zhang and Yang (2020).

The SVM, first introduced in 1995 by Cortes Zhang et al. (2020) is a classification algorithm that defines hyperplanes so as to maximize the margins, herein referred to as the distance between separating the hyperplane and the closest sample. The main advantage of SVM is attributed to its insensitivity to the amount of training data hence rendering it suitable for use in cases where there are limited training samples Foody and Mathur (2004). Ref Liu et al. (2006) employed SVM to perform forest disease classification on 1-meter resolution airborne images. Ref Van der Linden and Hostert (2009) adopted SVM to map land cover in urban areas using airborne imagery based on a resolution of 4 meters. A study in Asma and Abdelhamid (2020) proposed a novel approach for the classification of VHR remote sensing images by harmonizing the pixel-based and object-based classification techniques. The algorithm of super-pixels was employed to group pixels into different batches, usually referred to as segments. Super-pixels were then merged into more significant objects by using the metric distance between all neighbor segments. The resulting image was classified using Support Vector Machine into regions for water, trees, grass, and rocks. ML algorithms are also employed to detect changes e.g. forest change detection. For instance, Support Vector Machines (SVM) and genetic algorithms can be harmonized together to detect land cover changes. For this case, the radial basis kernel and associated parameters such as C and Ω for SVM are optimized using a genetic algorithm. This hybrid approach produced efficient results when implemented on the Mexico dataset and Sardinia dataset Pati et al. (2020). The challenge encountered in the SVM classifier is the selection of kernel parameters. The selection of parameters is done through a computationally intensive cross-validation process. The Radial Basis Function (RBF) based on the Gaussian function is the most widely used non-linear kernel

function in SVM. Selecting RBF is a challenging task since it involves defining appropriate range values for each parameter and determining the best combination through a cross-validation process. Another problem is that SVM-RBF's performance decreases whenever the number of features is much greater than the number of training samples.

Just like SVM, Rf is also a non-parametric classifier. The RF classifier is a bagging algorithm that uses a set of decision trees and classifies each instance based on the number of votes. RF is computationally efficient and is capable of handling high-dimensional data without over-fitting. Therefore, it has been successfully used in land cover mapping using VHR. Ref Adelabu et al. (2014) successfully used an RF classifier for insect defoliation classification, and Van Beijma et al. (2014) managed to classify forest habit on 2-meter resolution airborne imagery. A study in Cuypers et al. (2023) employed Random Forest Classifier on VHR optical imagery to improve object recognition for GEOBIA land use and land cover (LULC) classification. The study identified ten LULC classes on the satellite image obtained from Google Earth Engine in the city of Nice in France. The study investigated the impact of adding Gray-Level Co-occurrence Matrix (GLCM) texture information and spectral indices, and the results showed its classification accuracy from 67.05 to 74.30%. However, the RF classifier is very difficult to visualize and interpret in detail and it has proven to overfit for noisy datasets.

Deep neural networks are now getting much recognition in the field of semantic segmentation He, Zhang et al. (2016) Szegedy et al. (2017), image classification He, Zhang, et al. (2016) Szegedy et al. (2017), and object detection Redmon et al. (2016). This technology has quickly infiltrated remote sensing image applications, in particular, semantic segmentation classification has been widely used for land cover classification. With the aim of increasing accuracy in pixel-level land cover classification, a study in Dong et al. (2020) designed a feature ensemble network (FE-Net), comprising multi-scale feature encapsulation and two enhancement phases. The first phase adopts three layers which are shallow, middle, and deep-scale features from the ResNet-101 backbone and the second one is the multi-scale feature description enhancement. The optimal channel selection was also adopted to work on each intrascale and interscale feature sequentially. The model performed well as it achieved a classification accuracy of 68.08 and 65.16% on ISPRS and GID data sets respectively. In the same vein, a study in Zhou et al. (2023) proposed an EG-UNet enhancement model for open pit mining land cover with irregular and sparse spatial distribution features. The model is composed of two main modules, the edge feature enhancement module, and the long-range information extraction module. Since the edge of mine land contains more detailed information than other spectral locations, the Sobel operator was then used to extract object boundary, and this process gives an advantage of increasing the weight of features for preservation purposes before the pooling operation. The information extraction module's purpose is to extract tiny objects such as dumping grounds in the mining area. The EG-UNet model recorded the best performance, particularly on classifying classes with few samples. However, existing deep learning approaches in the area of remote sensing are still in their infancy and therefore lack a holistic approach Zhang et al. (2020). Also, deep learning models are black box in nature because of the complexity of their network structure such that it is very difficult to understand how they make decisions. Therefore, domain expert knowledge may not be certain if the model gained correct knowledge, hence undermining users' confidence in deep learning models Sarker (2021).

New methods such as Geographic Based Image Analysis (GEOBIA) have been of significant importance to the remote sensing science community. GEOBIA offers so many advantages over pixel-based methods. It can generate a large set of features by generating

more objects from the textural, spectral, and spatial properties of a group of pixels Souza-Filho et al. (2018). The ability of GEOBIA to process photos with a very high (spatial) resolution has led to its promotion as a tool for monitoring changes in agricultural, forested, and urban areas' land cover and land use Tompoulidou et al. (2016). The pixel-based approach ignored the fact that pixels are not isolated, but rather knitted together into a complex image with spectral patterns (Castilla and Hay 2008). It has been proven in VHR imagery that, individual pixels are too small to refer to a land cover class; therefore, they require a pixel footprint that is big enough to represent recurring elements such as forests (Blaschke and Strobl 2001). GEOBIA was introduced to provide answers to problems faced by pixel-based methods (Blaschke and Strobl 2001). GEOBIA is a branch of Geographic Information Science that aims to bridge the gap between the pixel and vector worlds (Castilla and Hay 2008; Blaschke 2010). GEOBIA, however, is heavily criticized for being excessively subjective because it approximates a degree of computer-aided photo interpretation Arvor et al. (2019). As a result of this, GEOBIA rules are not transferable, as their rules correspond to those of image processing chains. Therefore, GEOBIA is not suited to address issues related to the era of big data Arvor et al. (2013). Ontologies that provide a way of representing knowledge offer great potential to address such problems. They are able to represent numerical and symbolic knowledge, provide cognitive semantic reasoning capabilities, and exchange information on the deduced interpretation of remote sensing images. The definition of ontology derived from Artificial Intelligence is expressed as the formal, explicit specification of a shared conceptualization. From the definition, (1) formal means that the rules are expressed in a way that should be executed by computers, (2) explicit means that the definitions of all concepts and relations are clear and unambiguous, (3) shared means the definitions of all concepts and relations are commonly agreed by a community of knowledge domain. Formal ontologies provide shared definitions of concepts and associated relations to allow computer applications to communicate with each other Gruber (1995). They define the domain knowledge by expressing concepts and the relationships that bind them together (e.g. 'a woodland -is a kind of a –forest – – type', 'an orchard-is a kind of a-an – – artificial – – vegetation',etc.). Advantages brought by ontologies in remote sensing science applications with respect to description logic (DL) involve:

- Symbolic grounding. It precisely associates concepts with the right sensing data and also provides for valid associations of concepts between themselves. DL- ontologies represent low-level presentation into a high-level presentation which can be easily assimilated by human beings.
- Knowledge sharing. Standardization of ontology language and use of consensual conceptualization allows mechanisms for remote sensing image representations to be shared and reused by intelligent agents in the same domain.
- Reasoning. The description logic in ontology language provides a reasoning capacity that helps to infer new knowledge from existing explicit descriptions.

The rest of the paper is organized as follows. Sections 2 and 3 outline the current challenges with forest classifications using VHR images and the existing methods to handle VHR images, respectively. GEOBIA studies, challenges, and new developments are then considered in Sections 4–6, respectively. In Section 7, the use of ontologies for remote sensing image classification is explored. Section 8 proposes a state-of-the-art ontology-based model for forest image classification. Future directions and recommendations are highlighted in Sections 9 and 10 concludes the paper with a summary of findings.

2. Challenges with forest cover classification with VHR images

There are three major issues with using Remote Sensing (RS) pictures for forest cover categorization for VHR data, and these are as follows: (i) scaling up the well-trained classifiers from a single dataset leads to huge domain gaps across scenes and geographical locations; (ii) a lack of balanced, consistent, and high-quality training data hinders the development of accurate classifiers; and (iii) The impact of inter-class similarity and intra-class variability on classification accuracy.

2.1. Domain gaps across scenes and geographical locations

Transferability is a desired feature in trained models because data that would have been collected by different sensors in different geographical locations characterized by a variety of land patterns still achieve satisfactory results when compared with the actual training data. Applications related to computer vision have outstanding transferability, hence they are widely used in different domain setups Yosinski et al. (2014). Tasks such as semantic segmentation and tasks relating to the prediction of outdoor crowd-sourcing images generalize well because their prediction outcome is hugely determined by the structure of the scene when viewed from a ground view image Li and Snavely (2018). However, in RS images, the content of different parts of the images may vary greatly and thus are completely unstructured, and atmospheric effects create even greater variations in object appearances, let al.one the drastic change of land patterns across different geographical regions (e.g. urban vs. suburban, tropical vs. frigid). Therefore, transferability issues continue to be one of the main challenges to face when trying to scale up classification capabilities. In order to address this challenge, the Geometric-consistent Generative Adversarial Network (GcGAN) has been proposed by Fang et al. (2019) to eliminate any discrepancy that may arise between labeled and unlabeled images without losing their intrinsic land cover information by translating labeled feature images from the source domain to target domain. Another approach is the adoption of transfer learning models in remote sensing science applications, these techniques are able to produce a generalizable classifier by minimizing gaps in the feature space Qin and Liu (2022). These methods can be applied to data collected from a variety of sensors in a variety of geographical locations.

2.2. Lack of balanced, consistent, and high-quality training data

More training data is needed because both the amount of VHR data and the complexity of the models are growing. Traditional manual labeling methods, which were mostly used when processing coarse resolution data (like MODIS, Landsat, and Sentinel) Cai et al. (2014) or VHR data from a small area of interest (AOI), are not optimal and are no longer possible as the models are changed to DL models that need more data. To solve this problem, academics tried to get training data from many different places, such as crowd sourcing services (like Amazon Turk) [24] and public datasets (like OpenStreetMap) Haklay and Weber (2008). On the one hand, these extra datasets make it much easier to train high-accuracy classifiers, but on the other hand, they add new problems that may need to be solved for common training data problems that are explained below.

- Imbalanced training samples: When the classes or categories in the training data contain a varying number of images or samples, generally it causes the model to perform poorly in predictions. This was handled in traditional manual labeling approaches

because samples were drawn on purpose and reassembled afterward for shallow classifiers. However, for DL-based models, all available training data are often fed into a network, regardless of how balanced they are. In order to tackle the imbalance problem in VHR images Sun et al. (2020) developed an impartial semi-supervised learning approach based on extreme gradient boosting algorithm (ISS-XGB). The ISS-XGB incorporates several semi-supervised classifiers to solve the multi-class classification. The model first employs multi-group unlabeled data to suppress the imbalance of training samples and then uses extreme gradient boosting regression to simulate the target classes with positive and unlabeled samples.

- Inconsistent training samples: Researchers in the RS community who want to do semantic segmentation are now able to use more crowd-sourcing or public datasets Demir et al. (2018) Schmitt et al. (2021). But the class definitions and amount of detail in these crowd-sourcing datasets or public benchmark datasets may not be the same. Figure 1 shows an example of a more detailed classification that separates buildings from other man-made structures. Some datasets define the ground class as including low-vegetation, grass, and barren land, while others separate the ground into rangeland with low vegetation and barrens. So, the first problem with using this kind of data is figuring out how to change or improve their labels to fit specific needs and details about the classification jobs. Inconsistency is also caused by using different types of remote sensing data Sui et al. (2020) and also by having a data set with images consisting of different numbers of bands. Due to discrepancies in VHR data Jin et al. (2022) propose a multi-source data fusion technique that requires re-sampling to unify the spatial resolution. The technique filters training samples and has the ability to offer product correction at a fine resolution. The superpixel algorithm was adopted to correct unreliable information of multiple products into a new land cover

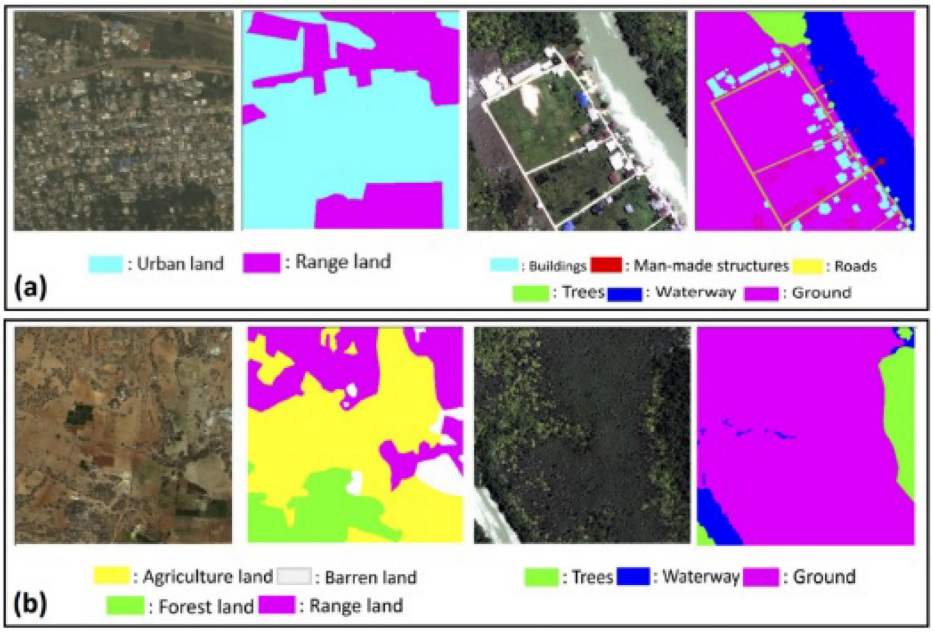


Figure 1. Inconsistencies of the class definition and level of details, (Qin and Liu 2022).

fusion product. The technique performed well as it achieved an accuracy of 85.80% on Landsat images.

- Lack of quality training data: The majority of present machine learning models in RS tend to underestimate their accuracy because they are contaminated by poor and low-quality training data, as shown in Schmitt et al. (2021). The low-quality data anticipated for learning algorithms can be another obstacle, despite active efforts to address the issue, for example by feeding the community with new data as samples. Employing techniques such as image enhancement and restoration helps to deal with issues to do with lack of quality data. Image enhancement techniques such as Histogram equalizer, Linear congruent adjustment, etc. improve image quality balancing parameters with regard to contrast, brightness, and sharpness Kundu (2022). Image restoration techniques artifacts like noise or blurs from images.

2.3. Intra-class variability and inter-class similarity for VHR data

VHR data with a ground resolution of a meter or less have improved earth observation by providing more detailed information. The increased resolution has increased intra-class variability and inter-class similarities: spectral information alone can identify a pixel as belonging to multiple land cover classes, and different classes may contain pixels with similar spectral signatures Qin (2015). There was extensive research into possible solutions for this problem, such as object-based approaches or spatial-spectral characteristics Ghamisi et al. (2015), but the advances that were possible couldn't keep up with the higher resolution and volume of data. Therefore, it may become more of a challenge when more sophisticated (and complex) models are utilized with larger numbers of annotated datasets. To tackle the problem of inter-class similarity and intra-class variance Venkataramanan et al. (2021) developed a model that automatically picks classes that are to be clustered and also determines an optimal number of classes to be generated. The obtained clusters are considered to be independent classes. Inter-class is dealt with by employing a triplet loss function to separate features between each class. Zhang et al. (2022) proposed a technique that tackles intra-class variance problems by developing a machine-learning model that organizes input instances as a graph. From the obtained graph, a normalized cut surrogate metric is used to determine intra-class variance within the training batch. The feature aggregation scheme is proposed by considering the equivalence between the normalized cut and random walk. The scheme is developed under the guidance of transition probability. Through supervision of aggregated features, transition probabilities are constrained to create a graph partition consistent with the given labels, hence the normalized cut and intra-class variance is well suppressed.

3. Proposed methods to handle VHR remote sensing images

The main obstacles to the current VHR RS picture classification are those already discussed. In addition to improving model performances, efforts have been made to solve these issues using multi-source/multi-resolution data, unlabeled data, more noise-tolerant models, and learning techniques. These initiatives can be generally described as follows:

- Weakly/Semi-supervised learning for small, imprecise, and incomplete samples.
  For weak supervision to work, the underlying training data must be inexpensive to collect (like publicly available GIS data) and noisy, imprecise, and asymmetrical. Because these methods attempt to incorporate heuristics, limitations, and error

distributions, that are unique to each situation, hence they are not universal. This is because semi-supervised learning in RS assumes the existence of a large amount of unlabeled data and relies on the limited training data to achieve high classification performance, and is thus relevant to applications that deal with crowd-sourcing labels or labels with minor temporal differences Larochelle (2020). Li et al. (2017) proposed the zero-short scene classification (ZSSC) that has got the ability to recognize images even when presented with incomplete labeled data. The model is attributed to its capacity to recognize images from unseen scene classes. The approach utilizes, word2vec, a natural language process to map names of seen/unseen scene classes to semantic vectors. The relationship that exists between seen and unseen classes are defined with the help of a semantic-directed graph constructed from the semantic vectors. To perform knowledge transfer from seen classes to unseen classes, an initial label prediction on a test image is performed, then the label propagation algorithm is developed for ZSSC. The label-refined approach is adopted to suppress noise in the zero-shot classification results. The approach outperformed the state-of-the-art learning models in scene classification.

- Transfer Learning and domain adaptation to fill domain gaps.
  Within the realm of machine learning, transfer learning (TL) is defined as the presumption that knowledge gained from completing one task may be valuable if transferred to completing another task. A model that learns to conduct per-pixel semantic segmentation of scenes, for instance, has the potential to make human detection more accurate. In the field of RS, this term mostly refers to methods that aim to produce a generalizable classifier by minimizing gaps in the feature space. These methods can be applied to data collected from a variety of sensors in a variety of geographical locations. Pan et al. (2016) designed a multi-layer transfer learning that caters to specific latent features for domain adaptation. Firstly, the model generates specific latent features, which are then combined together into one latent feature space layer. Since the layers have different pluralism, multiple layers are generated to correspond to each distribution layer. The difference in the pluralism in each layer means that learning distributions from one layer helps learn distributions on other layers. The iterative algorithm based on Non-Nagative Matrix Tri-Factorization was adopted to solve the optimization problem. The multi-layer transfer learning managed to outperform state-of-the-methods on sentiment classification tasks.
- Use low-resolution photos or public GIS data as sources of tagged or partially labeled data.
  Nearly 80% of the world's GIS data coverage is provided by OpenStreetMap, however, its quality varies Barrington-Leigh and Millard-Ball (2017), and some local governments make their GIS data available for public use. Researchers presented their work in this context and drew conclusions that were directly linked to the datasets. Additionally, these low-resolution labels can be used as a general guide to address domain gaps of data across various locations for scaling up the land cover classification of VHR data as the low-resolution labeled data with global coverage are gradually becoming more complete (e.g. National Land Cover Database Homer et al. (2012)). Wu et al. (2019) developed an effective unsupervised deep feature algorithm for classifying low-resolution images. The approach does not require any fine-tuning on the convenet filters and the convenet filters are used to extract features from both high and low-resolution images, and the obtained features are fed into a two-layer feature transfer network for knowledge transfer. The network has the ability to transfer distinguished features from a high-resolution feature space to a low-resolution feature space. The model was implemented on the VOC2007 dataset and showed significant improvement against baseline methods.

- Fusion of multi-modality and multi-view data. Unlabeled data sources such as Light Detection and Ranging (LiDAR), Synthetic Aperture Radar(SAR), and nighttime data can be used to study heuristics and improve latent representation learning Qin and Liu (2022). Lei et al. (2021) proposed a fusion of multi-modality and multi-scale attention network land cover classification of VHR images. The multi-modality fusion was designed on the basis of an encoding-decoding network that eliminates redundant features and fuses only useful features. This process increases the classification of land cover products by removing redundant features. The novel multi-scale spatial context enhancement module was adopted to improve feature fusion and alleviate the problem of large-scale variation of objects. The model was implemented on Vaihingen and Potsman datasets and performed well as it obtained F1-scores of 88.6 and 92.3% for Vaihingen and Potsman datasets, respectively.

3.1. Semisupervised learning (SSL) methods

One of the major challenges in Remote Sensing (RS) classifications is that the process of collecting VHR images for training (labeled) samples is really a tedious task. Therefore the RS science community has adopted SSL methods to tackle this challenge Yin et al. (2014); Bazi et al. (2012). SSL works by trying to generate a wealth of information from the available unlabeled data, despite having few available unlabeled data, with the aim of improving the performance of the classifier. Such approaches assume that points within the same structure are likely to have the same label Wang et al. (2015). In VHR images it seems reasonable to assume that if samples have close spectral information then they are likely to have similar labels. SSL methods have been successfully used in image classification applications such as vegetation mapping, land cover mapping, and urban planning Kwak and Kim (2023). Fan et al. (2020) proposed a semi-supervised multi-Convolutional Neural Network (CNN) ensemble learning method (Semi-MCNN) for urban land cover classification. The model harmonized the multi-CNN ensemble approach and a semi-supervised strategy to build an end-to-end architecture. This hybrid approach generally improves classification accuracy and generalization ability. The purpose semi-supervised technique was to leverage unlabeled images to labeled samples, and the ensembled teacher model dataset generation (EMDG), which is an automatic sample selection technique, was adopted to select appropriate samples and to generate large datasets from unlabeled samples automatically. The model was implemented on Shenzhen's land cover data and performed well as it achieved an overall accuracy of 92.5%. Ekanayake et al. (2018) developed a semi-supervised approach for mapping boundaries between two vegetation zones using satellite hyperspectral data. The approach employed the Maximum Likelihood Classification technique in order to detect pure vegetation pixels. In order to determine the boundary between two major vegetation zones, the technique considers the degree of correlation of pixels containing vegetation at various spatial coordinates. Finally, the systematic procedure comprising Fisher's Discriminant Analysis (FDA) and spectral clustering is used to divide the vegetation pixels into two vegetation zones.

3.2. Deep learning approaches

Deep learning technologies have been widely used to perform multi-class segmentation on VHR images Sertel et al. (2022). The number of classes to segment should be carefully examined prior to the application of deep learning technologies. Yuan et al. (2021) conducted a critical review on semantic segmentation using deep learning methods.

The findings from the study showed that segmentation of VHR on datasets such as ISPRS vaihingen (five classes), ISPRS Potsman (five classes), and Massachusets (two classes) achieved high accuracy ranging from 85 to 99%. Audebert et al. (2018) developed an efficient multi-scale deep fully CNN based on ResNet and SegNet with multi-modal to perform segmentation on high-resolution remote sensing data. Results obtained showed that the fusion of multi-modal data significantly increases the accuracy of semantic segmentation by attributing its capability to learn multi-modal features jointly. Fu et al. (2017) integrated Atrous convolution to Fully Convolution Network (FCN) to build multi-scale network architecture to perform semantic segmentation on VHR images obtained from GF-2 and IKONOS datasets. The Conditional Random Fields were also added to the network in order to refine the output class maps. The model performed well as it obtained the precision, recall, and kappa values of 0.81, 0.78, and 0.83, respectively. Other developments such as Densely Connected Convolutional Network (DenseNet) Huang et al. (2017), and ShuffleNet Zhang et al. (2018) have been extended in remote sensing segmentation to address issues to do computational complexity, and these designs are producing satisfactory performance in semantic segmentation for remote sensing data Chen, Fu et al. (2018). DenseNet is an extension of ResNet, which introduced extra connections from one layer to its subsequent layers, and this has increased information flow and feature reusing Huang et al. (2017). The building blocks of DenseNet are dense block, which is made up of stacked layers of two filters (a 3×3 followed by a 1×1 filter. The dense blocks are interconnected with a 1×1 convolutional layer for feature dimensionality reduction. The network structure alleviates the vanishing gradient problem and enables feature reuse. Figure 2 shows the DenseNet architecture.

ShuffleNet Zhang et al. (2018) significantly increases computational complexity by reducing computation complexity of 1×1 convolutions and utilizes channel shuffle to help the information flow across feature channels. The computation of individual shares to be processed by the GPU is divided by the group convolution, and the output is reorganized into a matrix, where the rows are the group count and the columns are the channel count. The depthwise convolution is used instead of 3×3 convolution. The second group convolution restores channel dimensionality to match the residual for concatenation. Figure 3 shows the DenseNet architecture.

3.3. Multi-resolution data classification

Multi-resolution data classification technique considers different levels of information granularity to analyze task data. This technique is widely used to perform classification

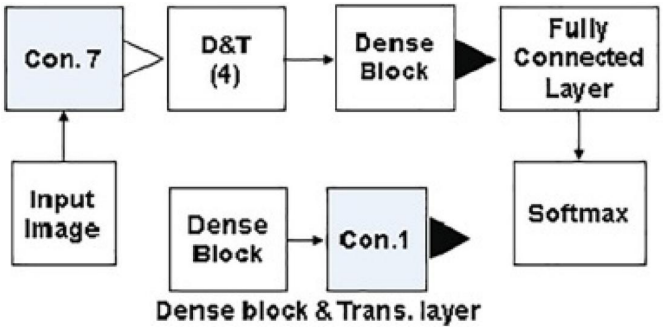


Figure 2. DenseNet architecture (Yuan et al. 2021).

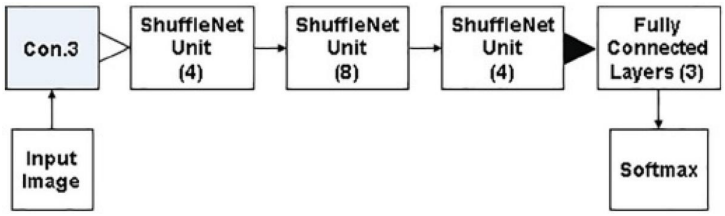


Figure 3. ShuffleNet architecture (Yuan et al. 2021).

tasks on data such as VHR images, graphs, and time series. It works by extracting patterns or features from images at different resolutions and integrating them into the classification performance. Duarte et al. (2018) developed a multi-resolution feature fusion for classifying building images using CNN. This approach integrates feature maps produced from different resolution levels (terrestrial, aerial, satellite) in order to categorize damages on building from remote sensing images. The results of the study demonstrated that multi-resolution fusion techniques outperform the traditional methods in classifying building images with 89% compared to 84%. The concept of using multi-resolution produces better accuracy and localization capability than using single-resolution features. Teruggi et al. (2020) proposed a hierarchical learning machine approach for multi-resolution 3D point cloud classification. The study extended the learning machine approaches with a multi-level and multi-resolution approach. The integration of the hierarchical concept optimized 3D classification results and improved the learning process. The multi-level and multi-resolution procedure was tested and assessed on two large datasets (the Pomposa Abby and Milan Cathedral both in Italy). The model managed to identify necessary architectural classes at each geometric resolution. Fixed network structures at a single resolution are difficult to characterize surface targets that have bright colors and different shapes with fixed sizes. To address this challenge Cong et al. (2022) proposed a structure defined by sample characteristic (SDSC) multi-resolution classification network that learns samples using a multi-resolution strategy and the principle of maximum classification probability. In order to improve the credibility and classification accuracy, the results obtained from the multi-resolution strategy were integrated into the final classification results. The proposed method is suitable for classifying high spatial resolution remote sensing images because of its better cognitive performance and insensitivity to noise.

4. GEOBIA studies

GEOBIA is a remote sensing tool used for land cover mapping and detecting land cover changes. It is a new discipline in remote sensing science that has evolved from pixel-based approaches and has significantly improved the workflow of imagery processing, particularly for land cover classification and detection Arvor et al. (2013). GEOBIA's main goal is to deal with more complicated classes that are determined by spatial and hierarchical relationships both within and outside of the classification process Lang (2008). Of course, one might perform a multi-spectral classification in an RS system first, then group and rearrange the labeled pixels to construct objects using GIS software. However, the analyst may be skewed by this sequence, which limits the number of classes that may be handled. The outcomes achieved through this process differ from those that would be obtained with a single conceptual step, as is the case with human perception. Instead of examining the spectral behavior of individual pixels, the object-based approach groups adjacent pixels into objects, which then serve as the observation units. This classification circumvents

the issue of artificially square objects as used in the per-pixel analysis Fisher (1997); Burnett and Blaschke (2003); Blaschke (2010), so long as the objects of interest cover a sufficient number of pixels to permit a meaningful representation of their shape. GEOBIA has been used in a range of applications such as geo-morphology Drăguț et al. (2011), agriculture Vogels et al. (2017), archeology research Hegyi et al. (2020), and soil science Dornik et al. (2018). GEOBIA has managed to bridge the gap between remote sensing and Geographic Information Science (GIScience). In the fraternity of GIScience, the term Object Image Analysis (OBIA) was first introduced in 2006 Lang and Blaschke (2006), and later reformulated as GEOBIA in 2008 whose central focus was on Earth Observation (EO) applications and the integration of geo-spatial-temporal reasoning to deal with high volumes of EO imagery and other related information extraction challenges Lang et al. (2019). GIScience scholars have reached a consensus on the fact that GEOBIA is a paradigm shift Blaschke et al. (2014), that has managed to bridge the semantic information gap from big data in the image domain. The representation of 2D imagery as a gridded array of pixels does not provide descriptive content with regard to semantic information or object boundaries Lang et al. (2019). Such information needs to be documented in metadata. The image content in the current setup cannot be queried, but attempts to meet this vision exist Blaschke (2010). Geographic Information system (GIS) datasets are discrete and the finite vector set handles discrete categorical nominal variables rather than numerical variables Lang et al. (2019). The success of GEOBIA as measured by bibliometric measures Blaschke et al. (2014) is attributed to its mediating power between geospatial entities and continuous field representations, which caters to the needs of GIS and remote sensing communities. The Harmonisation of these two models is presented in Figure 4.

The classification system of traditional pixel-based methods suffers from the salt and pepper-effect. This problem was alleviated by the Object-Based Image Analysis (OBIA) methodology when implemented on the Northern California vegetation inventory (Yu et al. 2006). OBIA adds object shape and context to spectral and textural information, and this significantly lowers the salt and pepper effect problem.

Another study by Chubey et al. (2006) used object-based analysis of IKONOS-2 imagery for extraction of forest inventory parameters rather than the traditional pixel-based image analysis approaches. Object-based analyses were first introduced in the area of remote sensing by Kettig and Landgrebe (1976), however, the approach did not receive much attention as its pixel-based method counterpart (Lu et al. 2013). Later, the object-based analysis techniques proved to be of significant importance in forest information extraction (Hay et al. 1996; Pekkarinen 2002; Imaging 2002). This was reinforced by the introduction of commercial object-based image analysis software such as eCognition (Arvor et al. 2019), feature analyst, etc. Chubey et al. (2006) developed a novel method that used

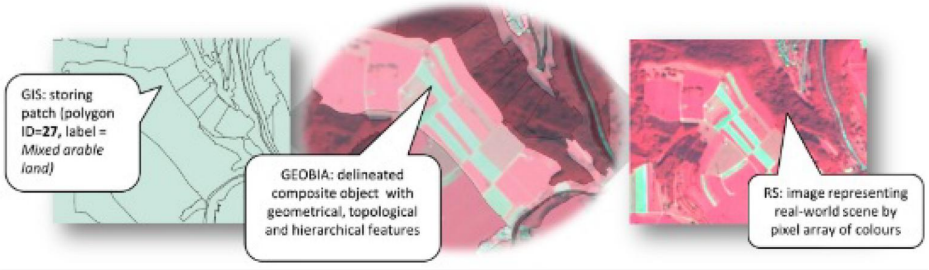


Figure 4. Impact of land cover type on the evolution of NPSS in the near-infrared (Lang et al. 2019).

the eCognition software and decision tree statistical analysis to extract forest inventory parameters. The IKONOS-2 images were segmented into image objects using the eCognition software. The multi-resolution segmentation was employed, where the image was partitioned into homogeneous multi-pixel regions. The size, spectral homogeneity, spatial homogeneity, and shape of the generated image objects were used to guide the segmentation procedure. The segmentation process was further tested against several other input/weighting combinations whereby each combination was evaluated on its ability to delineate meaningful landscape components. Image objects delineated carried crucial forest-related information that was derived from spectral and spatial characteristics of forest stand composition. Furthermore, analysis was performed using decision trees to determine correlations between image object metrics obtained from input data and individual forest inventory parameters. Decision trees were chosen because (a) they can handle high dimensional datasets, (b) they are able to work on both non-continuous and continuous variables, (c) of the non-parametric nature of the approach, and (d) they are easy to implement. However, there are some challenges with decision trees. Their performance depends on the quality and representation of training data (Friedl et al. 1999) and the accuracy increases with more training data. Therefore, the requirement for large training datasets is a concern from an operational perspective.

A quite number of studies have shown that OBIA methods are, re-applicable and more transferable to other images. This is achieved by re-applying the rule set on other conditions, and rule sets have the ability to adapt to new changed conditions. Hofmann et al. (2011) developed a new method to measure the robustness of a rule set. The new method is based on the assumption that the level of adaptation to be measured is in congruence with the quality of classification achieved. The robustness x_i of an unchanged rule set applied on an image M_i (i.e. $= Y = Y_I = Y_I$ is expressed by ratio of quality values: $x_i = \frac{q_i}{q_r}$. If $x_i > 1.0$ it implies a better result for M_i and vice versa for $x_i < 1.0$. The mean robustness of all the images M_n is expressed $x = \frac{1}{n}\sum_{i=r}^{n}(x_i)$ and the greater x the more Y for $q_r > 0$.

The studies in part delves into the importance of evaluating segmentation results and considered segmentation evaluation metrics such as the inverse of the number of objects (INO), Normalised Post Segmentation Standard Deviation, and Bhattacharyya Distance (BD) have been provided. A method for evaluating the quality of segmentation results in object-based classification was presented by Radoux and Defourny (2008). The proposed method constituted of two indices; one index was used to evaluate the extent to which the classification could be improved while the other assessed the boundary quality of the delineated land cover classes. Using a combination of three parameters from the same segmentation technique, the method was used to segment a Quick-bird image. It was established that large groups of pixels in an image, aid in the reduction of variance (Edwards and Cavalli-Sforza 1965). The study opted for a small intra-class variance with the assumption that it improves parametric classification. Over and under-segmentation were assessed using indices based on mean-sized objects. As a first quantitative goodness metric, the inverse of the Number of Objects (INO) was utilized. INO measures the ability of a model in segmenting an image into individual objects. INO is expressed as $INO = 1/N$ where N represents the number of objects. The second global index was the Normalised Post Segmentation Standard Deviation (NPSS). It examines segmentation quality based on the variability of the segmented image against the variability of the entire image. NPSS is expressed as $NPSS = (\sigma s - \sigma x)/\sigma x$ where σs is the standard deviation of the pixel intensity values in the segmented region and σx is the standard deviation of the intensity values of the whole image (that is it includes both the segmented region and

non-segmented region. The NPSS was used to calculate the class uniformity by replacing each pixel value with the parent object's mean values. The small intra-class variance does not always improve classification results; in some circumstances, a considerably large variance between two classes can improve the classification. Therefore, a dissimilarity metric, the Bhattacharyya Distance (BD) was co-opted in the study to test the relevance of the proposed goodness indices since it contains the term that compares co-variance matrices and it also accounts for classification errors (Webb 2003). BD is a measure of dissimilarity between two probability distributions. For probability distribution p and q on the same domain$X$, BD is expressed as $DB = -ln(BC(p,q))$ where $(BC(p,q)) = \sum_{x \in X} \sqrt{p(x)q(x)}$ where $p(x)$ and $q(x)$ are probability density functions.

Artifacts along the boundaries and missing boundaries are the key challenges with segmentation algorithms. The quality of segmentation precision is determined by the number of artifacts along the boundary. The accuracy and precision criteria proposed by Mowrer and Congalton (2000) were utilized to evaluate the positional quality of the edges. The bias and mean of the distribution of boundary errors were used to determine the accuracy and precision, respectively. Figure 5 shows sample errors along the edges of a segmentation result. Negative values were assigned to non-matching polygons (omission error) and positive values to matching cases. The goodness of indices was evaluated by NPSS and BD. Both indices gave valuable insight into segmentation findings. Results showed that NPSS was more correlated than INO. The positive results can be attributed to the fact that the mean class values were not modified by the segmentation. However, segmentation parameters were shown to be sensitive to global NPSS, with the object size parameter accounting for more than 80% of the variance. The effect of segmentation on every NPSS class had to vary, this reflects the sensitivity of segmentation algorithms to the land cover class. The absolute boundary error was sensitive to under-segmentation and was able to detect artifacts along class boundaries. There was a higher correlation $(R^2 > 0.94)$ between shape parameters and boundary errors. Results from Table 1 show the average absolute errors of parameters in various scales smooth, mixed, and compact. The studies have revealed that most segmentation algorithms face challenges that include artifacts and

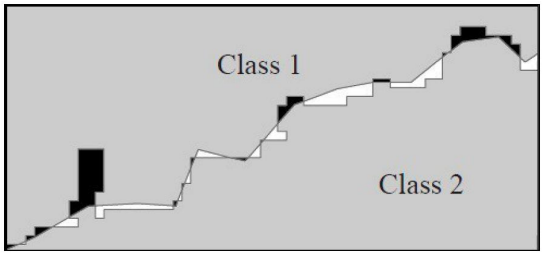


Figure 5. Errors along the edges of a segmentation output. Black polygons are omissions (−) and white polygons are commissions (+) with respect to class 1 (Xie et al. 2008).

Table 1. Average of absolute errors (in meters) on boundary position between deciduous and coniferous forests, for a combination of segmentation parameters, i.e. scale parameter between 10 and 60 and compactness, illustrated for forest/arable land interfaces (Xie et al. 2008).

Scale	10	20	30	40	50	50	Mean
Smooth	2.3	3.4	4.1	4.7	5.3	5.7	4.3
Mixed	2.2	3.2	3.8	4.3	4.8	5.3	3.9
Compact	2.1	3.0	3.7	4.2	4.7	5.1	3.8

missing values along the boundaries that deter them from achieving good segmentation results. Evaluation metrics such as NPSS, INO, BD, accuracy, and precision for assessing segmentation quality were looked at in these studies.

A study by Osio et al. (2018) uses OBIA-based monitoring of Riparian vegetation to assess the effect of flooding on the Lake Nakara Riparian Reserve vegetation species. An OBIA methodology was proposed (Osio et al. 2018) to serve as the basis for the classification of Riparian vegetation. The methodology comprised four pillars: data capture, pre-processing, processing, and analysis. Satellite data was downloaded from the USGS site, from Landsat 5 TM, Landsat 8 OLI (collected in 2014), and Landsat 8 OLI (collected in 2016) datasets. The pre-processing consisted of removing noise and ensuring uniformity between the datasets. The Ehlers fusion technique was employed to pen sharpen each image to 15 m resolution. Raw values of the images were converted to Top of the Atmosphere reflectance by the ArcGIS 10.4 software using a spatial analyst tool, in the arc toolbox. The planetary reflectance PY is defined as $PY = M_p Q_{cal} + A_p$, where Q_{cal} is the quantized and calibrated standard product pixel value and M_p and A_p are the band-specific multiplicative and additive re-scaling factors, respectively. A multi-resolution segmentation algorithm was adopted to convert pixels into image objects. Four bands namely; green, red, near-infrared (NIR), and shortwave infrared (SWIR) were used to classify vegetation indices on each dataset. NDVI values were obtained from the rule set established in the feature view and the supervised classification was carried out for each image using the K-NN algorithm. In terms of classification scales, scaling varied across different images such that there were different numbers of instances per imagery. Multi-resolution segmentation was used to segment images into image objects based on the feature parameters of layer weights, scale parameters, and composition of homogeneity criterion. The parameters were set in eCognition Developer 9.2 and were used by the multi-resolution segmentation to divide the image into homogeneous objects. Table 2 shows the segmentation scales set.

Hossain and Chen (2019) reviewed object-based image segmentation algorithms and challenges from remote sensing perspectives. The authors concluded that the quality of image segmentation has a significant impact on the final feature extraction and classification in OBIA (Hossain and Chen 2019; Vickers 2017; Su 2017). Many other studies (Blaschke et al. 2008; Cheng et al. 2001; Zhang et al. 2017) argued that the most crucial step in OBIA is image segmentation. Geographic Object Image Analysis (GEOBIA) was established to provide for image analysis by remote sensing scientists, environmental disciplines, and GIS specialists (De Jong and Van der Meer 2007). A comprehensive review of studies related to GEOBIA was undertaken by Blaschke (2010).

Chiu and Lin (2005) formulated the mathematical definition of segmentation as follows: given P, the homogeneity criteria and R, an entire image; R_i and R_j are segments of R if the following conditions hold (1) $R_i \subseteq R$, (2) $R = U_i = 1, \ldots, n$, (3) $R_i \cap R_j = \theta$ and (4) $P(R_i \cup R_j) = false$, where $i = j$ and R_i and R_j are neighbours. Segmentation algorithms have been categorized into (a) pixel-based (Friedl et al. 1999), (b) edge-based, (c) region-based, and (d) hybrid-based (Beveridge et al. 1989). In edge-based segmentation, the algorithm determines the edges, which are boundaries between objects (Cao et al. 2016); the edges are then closed up by continuous algorithms (Martin et al. 2004). Filtering,

Table 2. Segmentation scales (Osio et al. 2018).

Image	Scale	Shape	Compactness	Red	Blue	Green	NIR
Landsat5TM_2011	5	0.2	0.7	1	1	1	1
LandsatOLI_2014	40	0.5	0.5	1	1	1	1
LandsatOLI_2016	20	0.2	0.7	1	1	1	1

enhancement, and detection are the three key processes in edge detection (Jain et al. 1995). Filtering methods are necessary as they produce minimum blurring edges (Jain et al. 1995; Chen et al. 2006; Sahin and Ulusoy 2013). Enhancement highlights the pixels with huge changes in local intensity levels, and the enhanced data is utilized to detect real or genuine edges. The next stage is to use techniques like Hough transform (Kiryati et al. 1991), neighborhood search (Ghita and Whelan 2002), and watershed transformation (WT) (Vincent and Soille 1991). For natural segmentation, WT is commonly utilized (Hossain and Chen 2019). The region-based segmentation starts from the inside of the image and goes outwards until reaching the object boundaries (Zhang et al. 2016). Merging and splitting are the two basic operations in region-based segmentation (Fan et al. 2005). The segmentation process follows a systematic approach (Bins et al. 1996): (a) the first step performs an initial (seed) segmentation of the image, (b) the next step merges adjacent segments that are similar while splitting those that are dissimilar and (c) the previous step is repeated until there are no more segments to merge or split. The region growing or merging is defined by two main issues (Lucchese and Mitra 2001): (a) selection of a seed region and (b) similarity. Algorithms such as K-means clustering (Wang et al. 2010), hybrid region merging, single-seeded region growing (Verma et al. 2011), Particle Swarm Optimization (PSO) method, etc. (Mirghasemi et al. 2013) are used to generate the initial seeds. However, researchers are still in search of algorithms that work without seeds (Wu et al. 2015) or that are influenced by neighbors, even though seeded (Fan et al. 2005). After seed selection, the region grows sequentially by adding similar pixels, guided by specific homogeneity criteria. The criteria determine whether the pixels belong to the growing region or not (Nock and Nielsen 2004). The region splitting and merging entails, using the homogeneity criterion (based on attributes such as grey values, texture internal edges, etc.) to split the image into several segments (De Jong and Van der Meer 2007). If the seed image is not homogeneous, then the image is split into four sub-regions which serve as seeds for the next level (Martin et al. 2004). The process continues until all sub-regions become homogeneous. Bottom-up and top-down strategies are combined in the split and merge method (Guindon 1997). Bottom-up approaches enlarge the image by combining or merging comparable pixels, whereas top-down approaches split an entire image into image objects depending on the heterogeneity criterion (Benz et al. 2004). Edge-detecting methods face problems in generating closed segments and are excellent in detecting edges, while region-based methods are good in generating closed segments but are imprecise in detecting edges (Wang and Li 2014). Hence an algorithm was proposed that harmonized segmentation using both edge and region-based segmentation maps inputs Chu and Aggarwal (1993). The technique utilized the maximum likelihood estimator to predict initial edge positions from multiple inputs. An iterative procedure is then employed to smooth the resultant edge patterns. Finally, the edge map is converted to a region map using closed-edge contours. The regions are then merged to ensure that every region has the required properties.

5. Challenges of GEOBIA

In the past two decades, GEOBIA has been successfully adopted for land cover mapping (Blaschke and Strobl 2001; Blaschke et al. 2014). However, GEOBIA techniques require that regions of interest or objects be identified before applying classification rules on extracted objects (Blaschke and Strobl 2001). The segmentation step either relies on user expertise or empirical training to be adapted for each new scene to be processed (Drăguț et al. 2014; Ming et al. 2015). Hence, GEOBIA is not applicable for Big Geodata where

there is a large scale analysis which requires methods that are super-fast and robust (Merciol et al. 2018). Furthermore, GEOBIA has not yet been quantitatively verified though there is a general consensus among numerous researchers (Tehrany et al. 2014).

GEOBIA has been extensively used in land cover mapping applications. Land cover mapping is a complicated process as it incorporates factors such as image type, segmentation methods, accuracy assessment, classification algorithms, input features, etc. that have a great influence in the quality of the final product (Khatami et al. 2016). It is still a huge problem to come with a standard GEOBIA technique that provides an optimal solution for every study area. Spatial resolution is inversely proportional to segmentation scales. Figure 6 shows the relationship between spatial resolution and segmentation scale.

Whenever the spatial resolution becomes high, the segmentation scales become smaller and the lower the spatial resolution, the greater the configured optimal segmentation scales. It is very complex (Johnson and Xie 2011) to determine optimal segmentation scales due to the fact that the variability of the scale is affected by other image characteristics such as the size of the study area. The scale issue has emerged to be a huge problem for OBIA studies in relation to multi-segmentation scale methods. Therefore, there is a need to determine the appropriate segmentation scale necessary to obtain optimized segmentation results (Arbiol et al. 2007). Many researchers have explored trial and error approaches by varying segmentation scales based on their experience (Laliberte and Rango 2009), however, this approach is not advisable (Johnson and Xie 2011). In order to counteract this challenge Gu et al. (2018) propose an efficient multi-scale segmentation method based on graph theory and Fractal Net Evolution Approach (FNEA). The proposed model is shown in Figure 7. The contributions from this approach are that: (a) the Minimum Spanning Tree (MST) algorithm that performs the initial segmentation and the Minimum Heterogeneity Rule (MHR) algorithm adopted for object merging in FNEA are hybridized, (b) the segmentation strategy is implemented using data partition and the reverse searching forward processing chain using the Message Passing Interface (MPI) parallel technology. This approach is highly effective since it uses a fast graph segmentation algorithm and it also serves as a multi-scale segmentation and is hence suitable for a

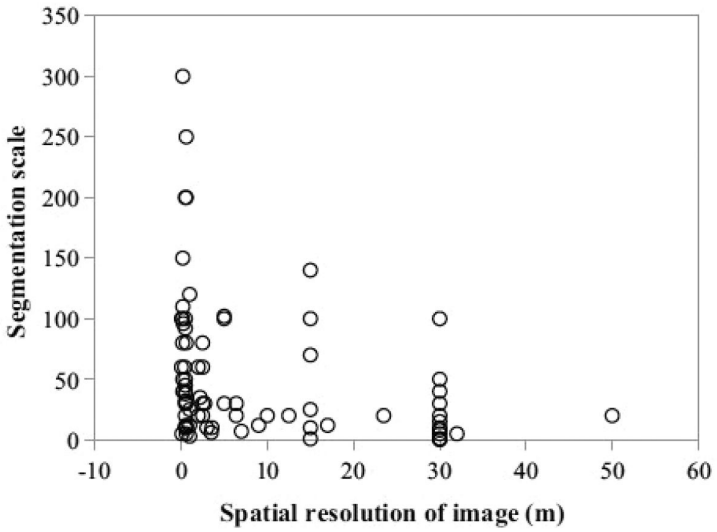


Figure 6. Correlation between spatial resolution and segmentation scale (Cai et al. 2014).

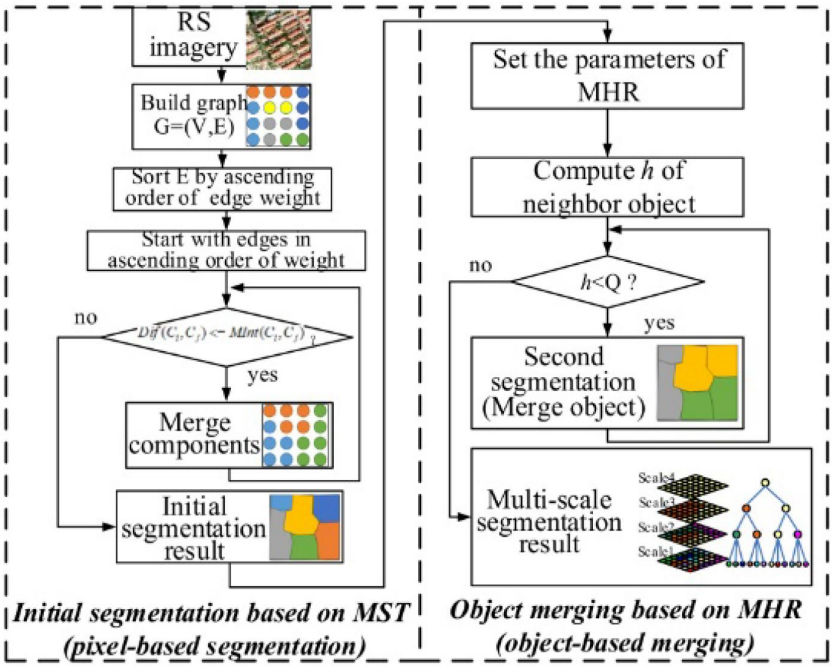


Figure 7. Multi-segmentation based on MST and MHR Gu et al. (2018).

variety of landscapes such as industrial or agriculture. The problem of multi-scale segmentation also arises in defining semantic rules to relate lower landscape units to high-level organizations. To address this issue Burnett and Blaschke (2003) developed Hierarchical Patch Dynamics (HPD) framework that aids in the development of describing patterns and processes, acting through a range of scales, which make up landscapes. The framework was implemented on two different projects. In the first project, habitat mapping was done using a multi-scale GIS database. The landscape segments were generated using sub-patch information including dominant tree crown densities and species. In the second project, fractal-based segmentation was adopted to produce agricultural scene segments, and the decision framework was adopted to choose the best combination of segmentation levels to identify shrub encroachment.

The next section presents recent developments in GEOBIA.

6. New GEOBIA developments

This section reviews new GEOBIA developments in terms of data sources, object based feature extraction, geo-object-based modelling frameworks, new forms of image objects, GEOBIA systems for novice GEOBIA users, and the use of knowledge from other disciplines.

6.1. Data sources

Modern high spatial resolution sensors provided a new landscape for remote sensing fraternity to study free-scale object or phenomenon from anywhere on the Earth's surface Chen, Weng, et al. (2018). Ancient GEOBIA studies used to work with classic, single-

image optical scenes for proof-of-concept studies, however new development in remote sensing fraternity have sharp increase of non-conventional data image type richer in spectral, spatial and temporal information, thereby, improving the modelling for geographical entities Chen, Weng, et al. (2018). Conventional GEOBIA data is defined as a high resolution imagery with limited spectral bands acquired by remote sensors mounted on relatively stable satellite/airborne platforms.

As depicted in Figure 8, rather than collecting images through satellite or airborne sensors, unmanned aerial system (UAS) or Drones have the ability to collect either sub-meter or sub-decimeter resolution data with high flexibility and very little demand for resources.

Similarly, Light Detection and Ranging (LiDAR) represents the conventional 2D spectral features with 3D structural information (bottom of Figure 8). Since segmentation in GEOBIA is solely applied to 2D imagery, LiDAR converts clouds or waveforms to raster format image models before they are used in GEOBIA framework. The GEOBIA community has taken the advantage of LiDAR's penetration capacity of retrieving 3D structures of non-solid objects with gaps such as trees. Chen, Weng, et al. (2018) argued that the LiDAR approach resembles real forest structure than sharpened WorldView-2 optical imagery at a 0.5 m resolution.

Hyperspectral images have been extensively used by GEOBIA experts to distinguish between geographical objects of similar spectral characteristics. Traditionally, hyperspectral images comprised of very limited spectral range that spans from visible to near-infrared section of the electromagnetic spectrum. These types of images have been successfully used in classifying mangrove species with 30-band (Kamal and Phinn 2011), examining post-fire severity by utilizing a 50-band MASTER mosaic (Powers et al. 2015) and assessing tropical forest area diversity with a 129-band AisaEAGLE imaging spectrometer (Schäfer et al. 2016). Hyperspectral imagery (middle of Figure 8) is very rich in spectral information as compared to multispectral imagery data sets, hence the extra bands can be used to obtain other useful information such as textural, object-based shape and contextual features.

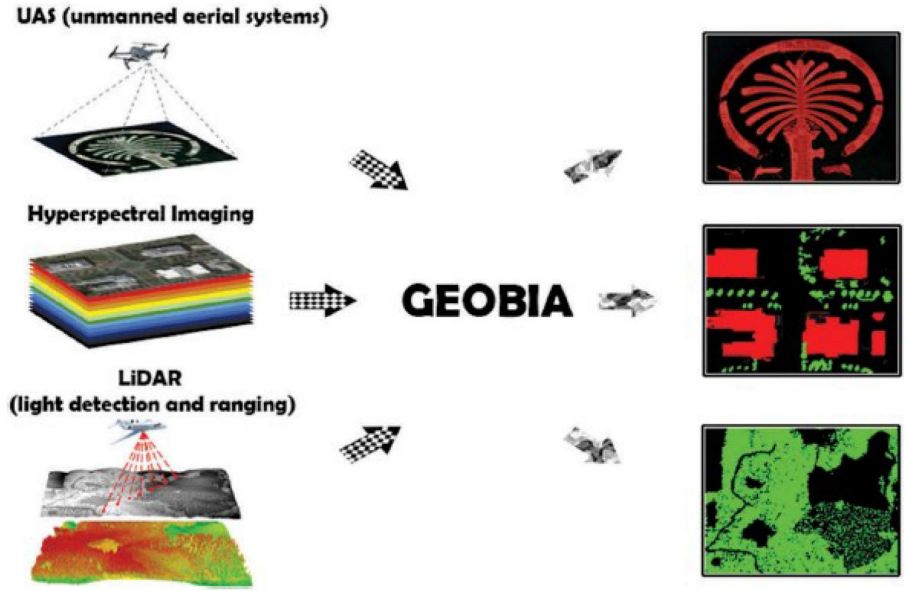


Figure 8. Conventional data types and image objects (Cai et al. 2014).

However, obtaining this extra information aids in computational costs due to rigorous analysis for feature extraction. Advances in sensor technology is moving towards extending spectral information beyond the red, green, blue and near-infrared segment of the spectrum, e.g. Worldview-2 has additional coastal-blue (400–450nm), yellow (585–625nm), red-edge (705–745nm) and near-infrared-2(840–1040nm) bands at the 1.84 m resolution Vermeulen and van Niekerk (2016).

6.2. Object based feature extraction

Classic features such as texture, context measures and spectral have been extracted through unsupervised image segmentation. GEOBIA has progressed to solely obtain these features through analyzing characteristics of geographic objects.

6.2.1. Novel object features

The traditional way of characterizing features of individual regions of interest involves measuring shape complexity (Mowrer and Congalton 2000; Cao et al. 2016), extracting features from interval-valued data modeling (He et al. 2016a), and establishing semivariogram descriptors for quantifying spatial correlation and patterns within objects as presented in Figure 9. A study by Wang et al. (2017) extended the approach to include the relationship between objects and lines. The authors explained that geographical objects in urban setups have more regular shapes than natural environments and have systematically distributed lines. Cai et al. (2014) extracted geostatistical features by examining the temporal behavior of each object's internal structure in relation to object based-based change detection.

A study by Wang et al. (2017) extended the approach to include the relationship between objects and lines. The authors explained that geographical objects in urban setups have more regular shapes than natural environments and have systematically distributed lines. Cai et al. (2014) extracted geostatistical features by examining the temporal behavior of each object's internal structure in relation to object based-based change detection.

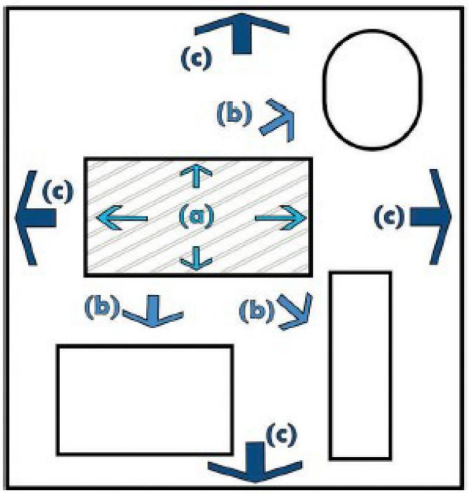


Figure 9. Object-based features obtained from three scales (a) image-objects (b) neighborhood image objects and (c) individual communities (Cai et al. 2014).

Chubey et al. (2006) proposed to incorporate neighboring image objects to improve the description of contextual information. Chen et al. (2011) came up with a geographic object-based image texture (GEOTEX) model that produced a set of texture measures by examining each image object with its corresponding neighbors through the natural window/kernel.

In GEOBIA techniques, the segmentation process produces a large number of object-based features (Figure 9), thereby reducing computational efficiency and also increasing modeling uncertainties. To mitigate this challenge, Powers et al. (2015) harmonized Principal Component Analysis (PCA) and Minimal Noise Reduction (MNR) to reduce the airborne MASTER sensor's 50 input spectral bands prior to segmentation.

6.2.2. Feature selection space

Techniques ranging from statistical analysis to machine learning and deep learning have been employed to obtain optimal features through the reduction of feature space. The majority of these techniques follow the approach of minimizing the number of input features while at the same time maximizing follow the separation distance between classes. Once processed features are then ranked in order of significance. Machine learning approaches have evolved for feature space reduction. However, for GEOBIA techniques, a consensus has not yet been reached as to which machine learning algorithm is the best for feature space reduction relative to particular applications. Algorithms that have been applied successfully for object-based feature selection include: Winnow (Littlestone 1988; Powers et al. 2015), random forest (Breiman 2001; Franklin et al. 2000), minimal redundancy maximum relevance (Peng et al. 2005) and Support Vector Machine (SVM) (Huang and Zhang 2013). Generally, the choice of algorithm for optimal feature space reduction depends on performance, ease of use, and accessibility. With the aim of improving feature selection from VHR images Chaib et al. (2022) proposed a framework based on Vision Transformer (ViT) models. Firstly, the ViT model is used to extract informative features from the VHR image scene, and the obtained features are merged into one signal dataset. The feature and selection algorithm is then adopted to trim off features that do not provide information to describe scenes such as beaches and agriculture. These features have a tendency of degrading the classification accuracy. The proposed model outperformed other state-of-the-art models when implemented on the VHR benchmark. Chen et al. (2016) proposed an efficient semi-supervised feature selection (ESFS) technique that selects all the desirable features by exploiting all the details available on the unlabeled objects. Firstly, it uses the probability matrix of unlabeled objects in the loss function to obtain features that are relevant per each class, instead of using traditional graphs. Lastly, norm regularization is employed to ensure that selection matrix rows have the required sparsity. ESFS outperformed other classical methods when implemented on a VHR image. Too and Abdullah (2021) proposed an improved genetic algorithm (GA) that incorporates the performance of GA in feature selection. The approach uses the competition strategy that integrates the new selection and cross-over schemes to improve the global search capability. Also, the dynamic mutation rate is also incorporated to improve the search power of the algorithm in the mutation process.

6.3. Geo-object based frameworks

GEOBIA has been dominantly used in land-cover/use classification. Recently it has also been successfully used to detect features in the area of archaeological remains (Lasaponara et al. 2016), green roofs (Theodoridou et al. 2017), alluvial fans (Pipaud and Lehmkuhl

2017), and dunes (Vaz et al. 2015). Various modifications have also been employed with the aim of meeting specific needs in real-world applications. Eckert et al. (2017) improved the classification of fine geographical objects which only existed in certain landscape zones. Another study by Guo et al. (2013) enrolled a two-step strategy to enhance classification by using object-neighbour context and scene context, respectively. However, these methods lack the ability to analyze latent spatial phenomenon. This challenge was addressed by Lang et al. (2008) who introduced geons which serve as a spatial units that are homogeneous in terms of varying space time phenomena under policy concern. Lang et al. (2014) further improved geons and came up with composite geons which provided solutions to policy-relevant phenomena such as societal vulnerability to hazards. The GEOBIA framework relied on geo-object based basic principles to derive image-object classes. However, improvement was done on GEOBIA workflow by adding parametric and non-parametric models to analyse object based variation within a specific class.

6.4. New forms of image objects

Considering the fact that the 3D geographical objects are represented as image objects in 2D format, some uncertainties and errors in identifying ground features could arise as some spatial dimensions are neglected (Alexander et al. 2010). Techniques in remote sensing and computer vision have progressed to the extent that it is now possible to capture geographical objects in 3D (Vaz et al. 2015). New trend in GEOBIA methods now incorporates vertical features in GEOBIA modelling e.g. carbon estimation at the tree cluster level using canopy height from a LiDAR sensor (Godwin et al. 2015). Extraction of object features by these techniques was centred on calculating boundaries of 2D image objects (Godwin et al. 2015; Zhang et al. 2013). Remote sensing science fraternity now advocates for generating image object features directly from a 3D scene model, which accurately represents real-world geographical objects. Photogrammetry and computer vision techniques can be used to construct 3D scene models (Luhmann et al. 2019). Using 3D information to delineate objects boundaries still remains a challenging task, as it is possible to loose crisp boundaries for certain components, for instance, a transition zone between wetland and water (Bian 2007). Defining fuzzy boundaries seemingly provides a better solution, whereby an image object is treated as homogeneous unit and equal parts (in terms of homogeneity) have a possibility of belonging to a certain class.

6.5. GEOBIA systems for novice GEOBIA users

GIScience and other communities took two decades to adopt GEOBIA frameworks and software packages (Chen, Weng, et al. 2018). The GEOBIA experts play a major role in incorporating user's knowledge and experience when developing GEOBIA models in order to achieve high accuracies. For GEOBIA applications to have a wider coverage, GEOBIA models should provide a way of translating novice-GEOBIA users understating of geographical entities into an appropriate choice of algorithms. This could be achieved with systems that have three key components such as (1) data query, (2) processing chain and (3) product sharing which can interactively direct a user to go through the entire GEOBIA process. Krizhevsky et al. (2012) argued that the translation of novice-GEOBIA understanding to GEOBIA language can be facilitated by the use of rule sets that are clearly defined and trained by machine or deep learning methods.

6.6. Embracing knowledge from other disciplines

GEOBIA has been extensively used in several disciplines (e.g. forest, urban planning, etc.), a gap still exists on how to integrate these disciplines to effectively support geo-object-based modelling (Chen, Weng, et al. 2018). Because of the nature of image analysis, GEOBIA hugely benefits from knowledge forecasting provided by computer vision that simulates human perception of digital imagery (Blaschke et al. 2014). Insufficient information on spectral, spatial, or temporal resolutions may result in experienced photo interpreters or computer programs producing incorrect perceptions of geographic entities (Castilla and Hay 2008). A potential solution proposed by Castilla and Hay (2008) is to take advantage of the Earth-centric nature of GEOBIA, where perceived geo-objects and their corresponding spatiotemporal dynamics meet rules or laws in natural or built-in environments. In a study on vegetation transitional zones from dense to bare ground in California by Chen, Weng, et al. (2018), a GEOBIA framework was employed to map disease-caused mortality and the results obtained indicated that there was over-estimation on patches of dead trees due to similar textural, geometrical and spectral characteristics between dead tree crowns and ground/shrub grass.

In order to enhance effective knowledge exchange and management, the GEOBIA community has adopted ontologies for specific applications (Andrés et al. 2017; Baraldi et al. 2017). Ontologies have played an important role in GEOBIA frameworks, but there is still a lack of comprehensive and universally accepted GEOBIA framework that provides guidelines for formalizing expert knowledge with ontologies. The next section discusses the use of ontologies in GEOBIA.

7. Ontologies for forest remote sensing image classification

Although data-driven approaches have attracted significant interest in research, knowledge-driven approaches remain an important future direction for the remote sensing (RS) science community Arvor et al. (2019). With that in regard, ontology having a strong power in knowledge representation, inference of common sense, knowledge sharing, and semantic cognition has gained much attention in the RS community Li et al. (2022). The Thailand Flora Ontology (TFO) (Panawong et al. 2018) was proposed to establish a semantic lexicon on the web, to assist plant biologists in the discovery of flora knowledge. The development of the ontology was against the backdrop that ordinary non-botanist people were not able to understand or receive accurate information about plants because plant information was expressed in English with botanical terms. Two steps were followed to design the ontology including the domain analysis for knowledge organization and the ontology development process. In the first step, a qualitative research method was employed to construct the Flora of Thailand knowledge structure, through (1) selection of already existing resources pertaining to plant ontology, biological classification, the flora of Thailand, and plant taxonomy, (2) flora content analysis from selected resources, (3) adoption of domain analytic approach for the organization of Thailand's Flora, and (4) explicit clarification of flora knowledge organization in consultation with domain experts. The ontology development process requires ontology engineers and developers to have the requisite knowledge in ontology specification and ontology development environment (Chansanam and Tuamsuk 2016). The construction of the TFO ontology followed the guidelines suggested in Ontology Development 101 by Noy and McGuinness (2001). The scope of the TFO ontology was limited to the Flora of Thailand, therefore the study recommended further development into an ontology-based retrieval system.

A study on ontology-based semantic mapping for integrating land cover products using hybrid ontology was presented by Zhu et al. (2021). The integration of land cover data depended on the characteristics of land cover products such as the thematic information, spatial resolution, temporal frequency and accuracy (Zhu et al. 2021). The integration was performed at the data and schema levels. The schema level integration used the Ontology-Based Data Integration (OBDI) approach. The OBDI has different variants: single-ontology schema-level, multiple-ontology, hybrid and Global-as-View ontology approaches (Ekaputra et al. 2017). The choice of the appropriate OBDI variant for land cover mapping and integration remains a key challenge. The study involved the use of multiple land cover products whose data sources had different semantics for land cover concepts. Therefore, the hybrid approach for ontology construction was adopted because of the heterogeneity of the data sources. Each land cover data source resulted in a distinct ontology. These local ontologies were subjected to a mapping process with the help of the EAGLE concept (Zhu et al. 2021). Figure 10 shows the OBDI structure. The global vocabulary construction was done by following the EAGLE matrix (Zhu et al. 2021). Firstly, the data source is refined to meet the specific requirements of the definitions of different land cover types. Thereafter, the characteristics that suit a given cover type are arbitrarily chosen for the construction of local ontologies. The land cover type integration is facilitated by adding the axioms and attributes to reduce design inconsistencies. The conceptual description of each data source is explicitly done by local ontologies. Terms of interest of each data source and the hierarchical relationships between classes are analyzed. Then, local ontologies of land cover concepts are disintegrated to express the attributes and relationship clearly.

Figure 10 shows the three building blocks of the EAGLE matrix, that is, the land cover component (LCCs), land use attributes (LUA), and further characteristics (CH). Moving down the matrix, there is grain granularity, whereby the grain is refined going down in order to meet the requirements of the definition of the different scales of land cover types. The ontology of a particular land cover product can be designed by choosing an appropriate combination of components, attributes, and characteristics from the EAGLE matrix. The EAGLE matrix is extended to describe the main components and the relationship between them. Figure 11 shows the architecture of the EAGLE matrix. The conceptual model of the

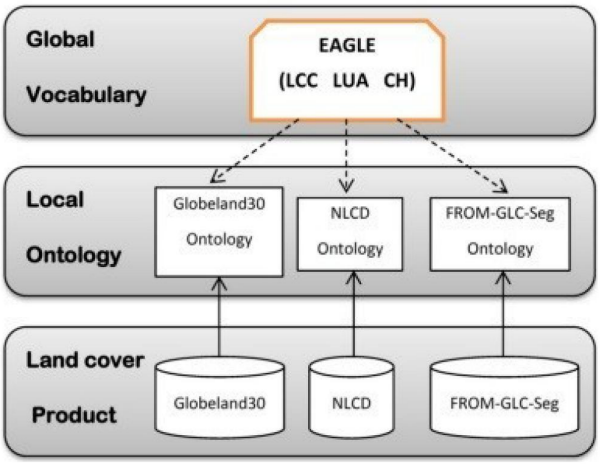


Figure 10. Ontology construction diagram (Zhu et al. 2021).

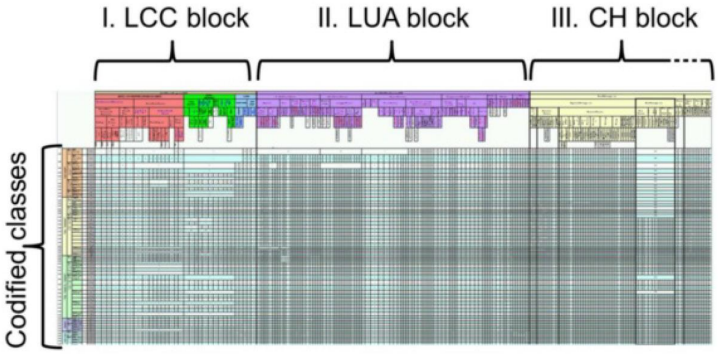


Figure 11. Structure of EAGLE matrix showing three blocks of land cover components (LCC), land use attributes (LUA), and characteristics (CH) (Smith and Hazeu 2015).

data source is described using a local ontology. All the required data sources are carefully examined by assessing the terms of each data source and the hierarchical relationship between each class. Finally, the ontology cover of each concept is broken into the global vocabulary-EAGLE matrix to clearly express the relationship between attributes. The land products considered in this example include NLCD, Glodeland30, and FROM-GLS-seg. The classes in the land cover product are organized in a hierarchical structure. Figure 12 shows an example of a coniferous forest in the local ontology of FROM-GLC-seg.

A study in Li et al. (2022) proposed a collaborative boosting framework(CBF) that integrates a deep learning approach and a knowledge-driven ontology reasoning module for remote sensing image semantic segmentation. The approach consists of two main modules, that is, the deep learning module based on the semantic segmentation network (DSSN) and the ontology reasoning module. The ontology reasoning module's role is to establish a connection between intra- and extra-taxonomy reasoning in series. The intra-taxonomy reasoning module is incorporated to correct misclassifications done by the DSSN thereby improving the interpretability of the classification. On the other hand, the extra-taxonomy reasoning module works on the corrected results in order to provide refined details that will help DSSN make reliable interpretations. The two modules in the model interact iteratively until the output from the entire system is optimized. The CBF model's primary focus is on predicting information relating to elevation and shadow on the basis that the predictions by the extra-taxonomy reasoning are sufficiently accurate for the DSSN.

8. State of the ontology-based model for forest image classification

A state-of-the-art model based on ontology and a deep learning model in Kwenda et al. (2023) was employed to classify forest images into their respective categories. The basis of this study was derived from the notion that integrating ontologies and semantic relationships significantly increases image classification accuracy. The model is composed of three main phases, that is (1) feature extraction, (2) ontology building, and (3) image classification. The model is presented in Figure 13.

8.1. Feature extraction

In image processing tasks features play a critical role in image classification. The ensemble of ResNet50, VGG16, and Xception deep learning approaches was used to generate a set

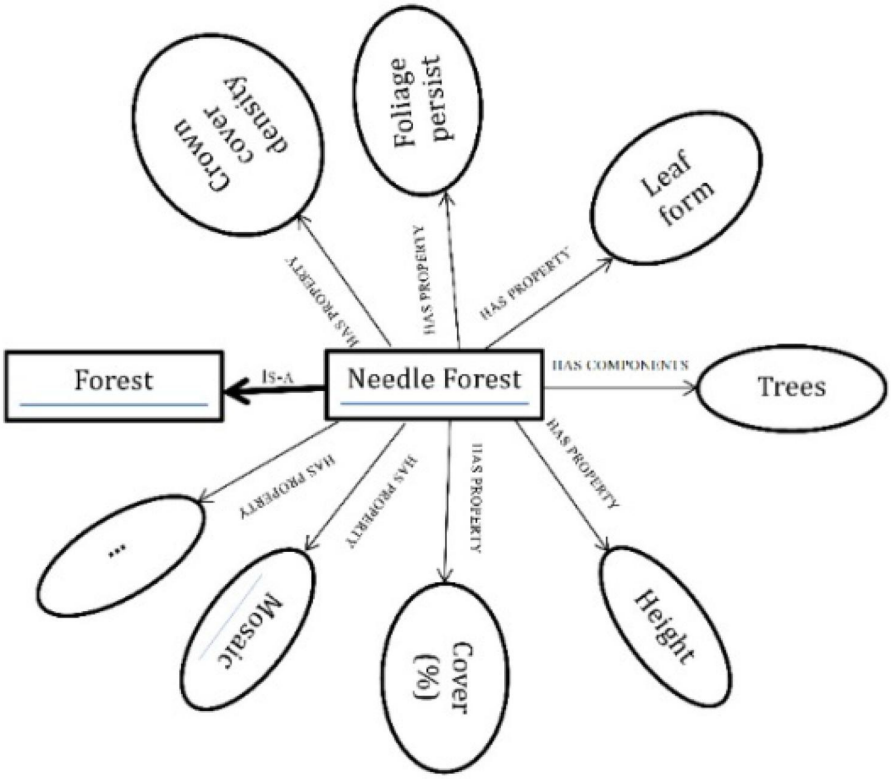


Figure 12. Coniferous forest in the local ontology of from-GLC-Seg. (Zhu et al. 2021).

of features from the training data set. Features produced by each deep learning technique were aggregated together to produce the final feature vector. The ensemble approach generates features that produce more accurate results than those generated by a single approach.

8.2. Ontology building

The process of building ontology was synthesized through concept extraction and relation generation. Concepts relating to forests were established and the associated relationships between the concepts were generated. The semantic relationship between image classes helps train images for classes and this is accomplished by grouping together images belonging to a particular class. Image that belongs to a particular class C_i denoted as x_i implies that x_i is a child of C_i. Suppose that 'artificial crop vegetation' and 'natural crop vegetation' are the superclasses at the root node, the semantic rules will categorize images of 'field' and 'orchard' to 'natural vegetation node' even though both nodes belong to the 'Primary Vegetative Area' parent class (Figure 14). The relationship between two concepts is shown with a relationship arrow that joins the concepts together, e.g. an arrow from 'orchard' concept to 'Artificial crop Vegetation Concept' implies that 'orchard' is_ a 'Artificial crop Vegetation'. The type of relationships considered in the study were hyponymy and hypernymy relationships. The generated ontology is shown in Figure 14.

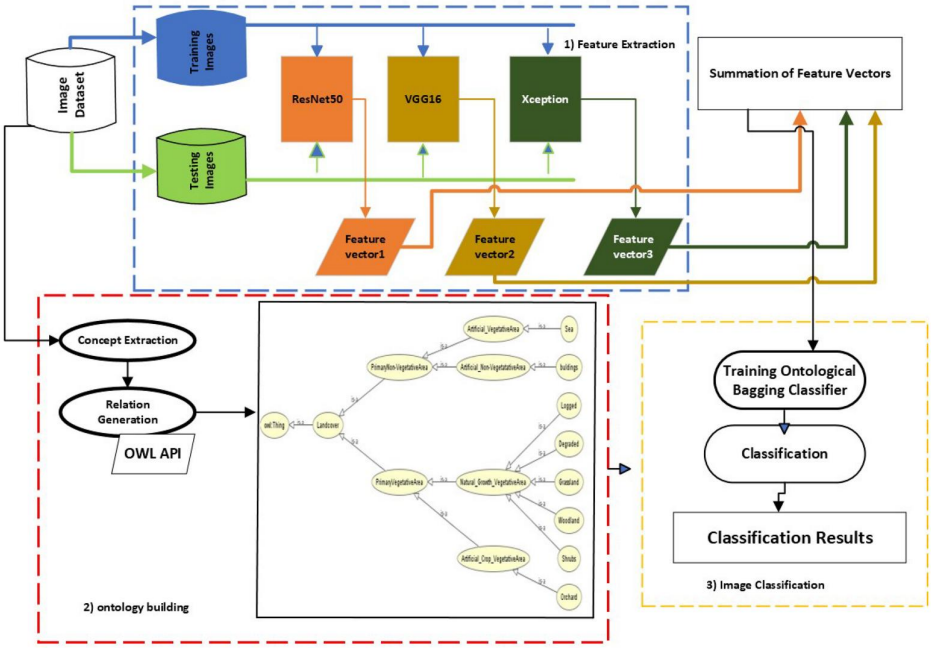


Figure 13. The proposed model.

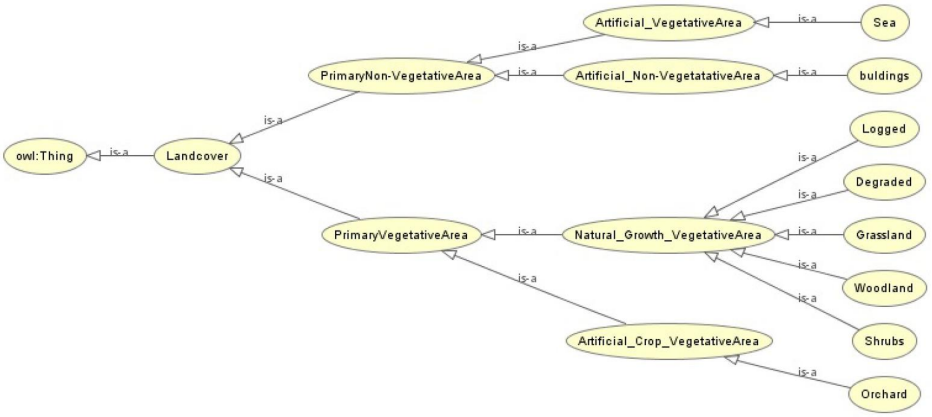


Figure 14. Ontology of Forest types Kwenda et al. (2023).

8.3. Image classification

The set of features generated by the ensemble of deep learning approaches was used to train a one-vs-all Support Vector Machine (SVM) classifier for each classifier so that each class could be distinguished from other classes. The classification of a given test image is performed by both hyponymy and hypernymy classifiers as illustrated in Figure 15. A given test image is assigned to a class with the best hypernymy classifier (artificial_ crop_ vegetation and the best

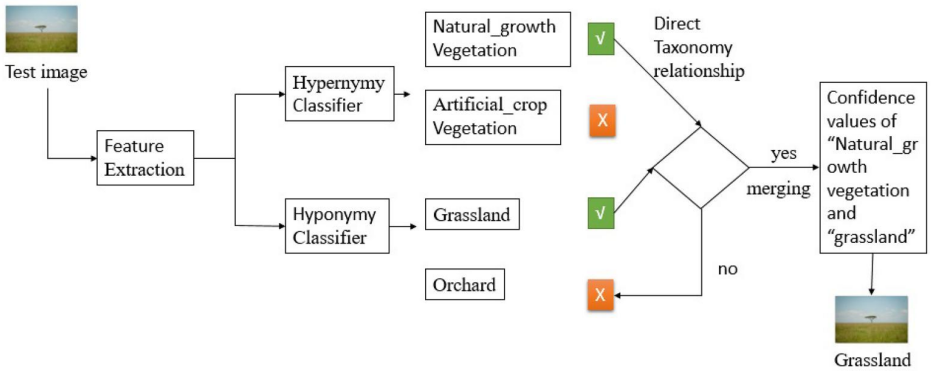


Figure 15. Classification using merging classifiers Kwenda et al. (2023).

hyponymy classifier (grassland). If the classifiers have a direct relationship, then their output will be merged together else the hyponymy classifier will be considered.

8.4. Evaluation metrics for the state-of-the-Art model

Metrics such as confusion matrix, Root Mean Square Error (RMSE), Accuracy, and Receiver Operating Characteristics Area Under the Curve (ROC_ AUC) were used to evaluate the model. Accuracy returns a ratio of correctly predicted classes to the number of samples evaluated. The definition of accuracy is presented in Equation (1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Where *TP, TN, FP*, and *FN* represent true positives, true negatives, false positives, and false negatives respectively. The area under the ROC curve is generally used to assess comparisons between learning algorithms, as well as the establishment of an optimal learning model. The AUC values rank the performance of the classifier. The AUC is presented in Equation (2).

$$AUC = \frac{s_p - n_p(n_n + 1)/2}{n_p n_n} \tag{2}$$

Where s_p is the sum of positive ranked samples, n_n, n_p denotes the number of negative and positive samples respectively. RMSE returns the square root mean square of all errors. The RMSE is expressed as Equation 3.

$$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^{n} (O_i - P_i)^2}. \tag{3}$$

where O_i are the actual values and P_i are the predicted values. With a confusion matrix, it becomes very easy to identify classes with more mislabelled data than others by providing the visualization performance of the classifiers.

Table 3. Quantitative comparison of models.

Model	Test images	Correctly classified	misclassified	ROC_ AUC	RMSE	Accuracy
kNN	152	124	28	0.97	1.530	0.816
OntologicalBagging	152	146	6	0.99	0.532	0.961
RF	152	131	21	0.99	1.094	0.862
DecisionTree	152	98	54	0.81	2.090	0.645
SVM	152	135	17	0.98	1.048	0.888
GaussianNB	152	97	55	0.79	1.678	0.638

8.5. Results of the state of the art ontology-based model

Results shown in Table 3 indicate that the state-of-the-art model based on ontology outperformed baseline classifiers without ontology in terms of ROC_ AUC, RMSE, and Accuracy. The ontology-based model performed well in separating classes in relation to other models as it attained the highest ROC_ AUC value of 0.99. The model also recorded the lowest RMSE of 0.532 suggesting that the predictions made by the model were very close to the actual values.

9. Future directions and recommendations

A study in Arvor et al. (2019) recommended two research directions for using ontologies in remote sensing science namely, (a) the modeling of ontologies with spatial reasoning and cognitive semantics and (b) the investigation of bottom-up vs top-down approaches. The study also advised that spatiotemporal information can be included in ontologies by adopting the principles of naive geography and cognitive geography during the development process. In fact, these principles reflect two primary roles of ontologies that are, the alignment of data with expert knowledge and the representation of common sense categorization based on expert conceptualization. The challenges of spatiotemporal ontologies to implement cognitive semantics was emphasized by Kuhn et al. (2007). The authors provided guidelines for developing geospatial ontologies that are more cognitive, including (a) the use of sound meaningful and suitable primitives, (b) the recognition of space and time as the foundational aspects of ontology because they correlate with human conceptualization, (c) use process-oriented rather than static structures, (d) harmonize realistic semantics and cognitive semantics, (e) allow perspectivism and relativism, (f) allow conceptual mapping to enhance human-computer interaction, and (g) consider contextualization of elements to relate the situational and individual settings. Though GEOBIA came as a relief to the remote sensing science community by providing solutions to the problems posed by pixel-based methods, its use in forest image analysis suffers a setback of localizing knowledge within a particular domain which is rarely transferable because it solely depends on expert knowledge Thus, forest image analysis and classification require expert knowledge from different facets of remote sensing professionals, which is rarely formalized and difficult to automate, The study recommends the adoption of ontologies for forest image analysis and classification as they promote knowledge sharing and the reuse of formalized remote sensing expert knowledge. This recommendation is supported in study Arvor et al. (2019) that ontologies in remote sensing provide a breakthrough in dealing with the complex definition of geographic concepts, handling a geographic concept's vagueness and ambiguity, and managing sensory and semantic gaps. We recommend a hybridization approach of ontologies and Explainable Artificial intelligence (XAI) for image classification in remote sensing science thereby boosting domain expert confidence in the results produced. Such a hybrid approach provides an explicit explanation of

how it reaches its conclusion. Production of features is to be performed by XAI, and classifiers trained through ontology are to perform the classification task.

10. Conclusion

This paper conducted a critical survey of GEOBIA methods for forest image detection and classification. A review of modern ontology-based remote sensing applications for forest image classification gave an insight into the power of ontologies to explicitly represent knowledge, thereby, improving the classification process. The shortcomings of GEOBIA such as failure to deal with segmentation scales, highly subjective because of computer-aided photo interpretation, and not being able to handle Big Geodata information were addressed, and the call for ontologies in remote sensing applications as a solution for GEOBIA problems was highlighted. The primary core of representing domain expert expression has attracted the adoption of ontologies in remote sensing applications. The study recommended the revamp of GEOBIA, by adopting a hybridization approach of XAI and ontological frameworks. Considering that XIA is not a black box in nature and it provides an explicit explanation of how they reach their conclusion, the study recommended the approach to be used for feature generation. In this regard, the domain expert's confidence in the obtained results is raised. Feature vector from XAI is passed on to classifiers trained through an ontological framework to perform the final step of segmentation.

Acknowledgements

The authors thank the University of KwaZulu Natal for providing financial assistance in accessing all resources and tools required to undertake this study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

References

Adelabu S, Mutanga O, Adam E. 2014. Evaluating the impact of red-edge band from rapideye image for classifying insect defoliation levels. ISPRS J Photogramm Remote Sens. 95:34–41. doi: 10.1016/j.isprsjprs.2014.05.013.

Alexander C, Tansey K, Kaduk J, Holland D, Tate NJ. 2010. Backscatter coefficient as an attribute for the classification of full-waveform airborne laser scanning data in urban areas. ISPRS J Photogramm Remote Sens. 65(5):423–432. doi: 10.1016/j.isprsjprs.2010.05.002.

Andrés S, Arvor D, Mougenot I, Libourel T, Durieux L. 2017. Ontology-based classification of remote sensing images using spectral rules. Computers Geosci. 102:158–166. doi: 10.1016/j.cageo.2017.02.018.

Arbiol R, Zhang Y, I Comellas VP. 2007. Advanced classification techniques: a review. Revista Catalana de Geografia.

Arvor D, Belgiu M, Falomir Z, Mougenot I, Durieux L. 2019. Ontologies to interpret remote sensing images: why do we need them? GIScience Remote Sens. 56(6):911–939. doi: 10.1080/15481603.2019.1587890.

Arvor D, Durieux L, Andrés S, Laporte MA. 2013. Advances in geographic object-based image analysis with ontologies: a review of main contributions and limitations from a remote sensing perspective. ISPRS J Photogramm Remote Sens. 82:125–137. doi: 10.1016/j.isprsjprs.2013.05.003.

Asma SB, Abdelhamid D. 2020. An object-based approach to VHR image classification. In 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS), Tunis, in Tunisia; IEEE. p. 93–96. doi: 10.1109/M2GARSS47143.2020.9105140.

Audebert N, Le Saux B, Lefèvre S. 2018. Beyond RGB: very high resolution urban remote sensing with multimodal deep networks. ISPRS J Photogramm Remote Sens. 140:20–32. doi: 10.1016/j.isprsjprs.2017.11.011.

Baraldi A, Tiede D, Sudmanns M, Belgiu M, Lang S. 2017. Systematic ESA EO level 2 product generation as pre-condition to semantic content-based image retrieval and information/knowledge discovery in EO image databases. In Proceedings of the 2017 Conference on Big Data from Space; Luxembourg: Publications Office of the European Union Luxembourg. p. 17–20.

Barrington-Leigh C, Millard-Ball A. 2017. The world's user-generated road map is more than 80% complete. PLOS One. 12(8):e0180698. doi: 10.1371/journal.pone.0180698.

Bazi Y, Alajlan N, Melgani F. 2012. Improved estimation of water chlorophyll concentration with semisupervised gaussian process regression. IEEE Trans Geosci Remote Sensing. 50(7):2733–2743. doi: 10.1109/TGRS.2011.2174246.

Benz UC, Hofmann P, Willhauck G, Lingenfelder I, Heynen M. 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS J Photogramm Remote Sens. 58(3–4):239–258. doi: 10.1016/j.isprsjprs.2003.10.002.

Beveridge JR, Griffith J, Kohler RR, Hanson AR, Riseman EM. 1989. Segmenting images using localized histograms and region merging. Int J Comput Vision. 2(3):311–347. doi: 10.1007/BF00158168.

Bian L. 2007. Object-oriented representation of environmental phenomena: is everything best represented as an object? Ann Assoc Am Geograph 97(2):267–281. doi: 10.1111/j.1467-8306.2007.00535.x.

Bins LS, Fonseca LG, Erthal GJ, Ii FM. 1996. Satellite imagery segmentation: a region growing approach. Simpósio Brasileiro de Sensoriamento Remoto. 8(1996):677–680.

Blaschke T. 2010. Object based image analysis for remote sensing. ISPRS J Photogramm Remote Sens. 65(1):2–16. doi: 10.1016/j.isprsjprs.2009.06.004.

Blaschke T, Hay GJ, Kelly M, Lang S, Hofmann P, Addink E, Feitosa RQ, Van der Meer F, Van der Werff H, Van Coillie F, et al. 2014. Geographic object-based image analysis–towards a new paradigm. ISPRS J Photogramm Remote Sens. 87(100):180–191. doi: 10.1016/j.isprsjprs.2013.09.014.

Blaschke T, Lang S, Hay G. 2008. Object-based image analysis: spatial concepts for knowledge-driven remote sensing applications. Dordrecht, Netherlands: Springer Science & Business Media.

Blaschke T, Strobl J. 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and gis. Zeitschrift für Geoinformationssysteme. 14(6):12–17.

Breiman L. 2001. Random forests. Mach Learn. 45(1):5–32. doi: 10.1023/A:1010933404324.

Burnett C, Blaschke T. 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. Ecol Modell. 168(3):233–249. doi: 10.1016/S0304-3800(03)00139-X.

Cai S, Liu D, Sulla-Menashe D, Friedl MA. 2014. Enhancing Modis land cover product with a spatial–temporal modeling algorithm. Remote Sens Environ. 147:243–255. doi: 10.1016/j.rse.2014.03.012.

Cao W, Li J, Liu J, Zhang P. 2016. Two improved segmentation algorithms for whole cardiac ct sequence images. In 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Datong, China; IEEE. p. 346–351. doi: 10.1109/CISP-BMEI.2016.7852734.

Castilla G, Hay G. 2008. Image objects and geographic objects. In: Blaschke T, Lang S, Hay GJ, editors. Object-based image analysis. Lecture notes in geoinformation and cartography. Berlin, Heidelberg: Springer; p. 91–110.

Chaib S, Mansouri DEK, Omara I, Hagag A, Dhelim S, Bensaber DA. 2022. On the co-selection of vision transformer features and images for very high-resolution image scene classification. Remote Sensing. 14(22):5817. doi: 10.3390/rs14225817.

Chansanam W, Tuamsuk K. 2016. Development of imaginary beings ontology. In International Conference on Asian Digital Libraries, Tsukuba, Japan; Springer. p. 231–242.

Chen G, Hay GJ, Castilla G, St-Onge B, Powers R. 2011. A multiscale geographic object-based image analysis to estimate Lidar-measured forest canopy height using quickbird imagery. Int J Geograph Inform Sci. 25(6):877–893. doi: 10.1080/13658816.2010.496729.

Chen G, Weng Q, Hay G, He Y. 2018. Geographic object-based image analysis (Geobia): emerging trends and future opportunities. Giscience Remote Sens. 55(2):159–182. doi: 10.1080/15481603.2018.1426092.

Chen K, Fu K, Yan M, Gao X, Sun X, Wei X. 2018. Semantic segmentation of aerial images with shuffling convolutional neural networks. IEEE Geosci Remote Sensing Lett. 15(2):173–177. doi: 10.1109/LGRS.2017.2778181.

Chen X, Song L, Hou Y, Shao G. 2016. Efficient semi-supervised feature selection for VHR remote sensing images. In 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China. IEEE. p. 1500–1503.

Chen Z, Zhao Z, Gong P, Zeng B. 2006. A new process for the segmentation of high resolution remote sensing imagery. Int J Remote Sens. 27(22):4991–5001. doi: 10.1080/01431160600658131.

Cheng HD, Jiang XH, Sun Y, Wang J. 2001. Color image segmentation: advances and prospects. Pattern Recognit. 34(12):2259–2281. doi: 10.1016/S0031-3203(00)00149-7.

Chiu KY, Lin SF. 2005. Lane detection using color-based segmentation. In: IEEE Proceedings. Intelligent Vehicles Symposium, Las Vegas, Nevada, USA; 2005; IEEE. p. 706–711.

Chu CC, Aggarwal JK. 1993. The integration of image segmentation maps using region and edge information.IEEE Trans Pattern Anal Machine Intell. 15(12):1241–1252. doi: 10.1109/34.250843.

Chubey MS, Franklin SE, Wulder MA. 2006. Object-based analysis of ikonos-2 imagery for extraction of forest inventory parameters. Photogramm Eng Remote Sens. 72(4):383–394. doi: 10.14358/PERS.72.4.383.

Cong M, Xi J, Han L, Gu J, Yang L, Tao Y, Xu M. 2022. Multi-resolution classification network for high-resolution UAV remote sensing images. Geocarto Int. 37(11):3116–3140. doi: 10.1080/10106049.2020.1852614.

Cortes C, Vapnik V. 1995. Support-vector networks. Mach Learn. 20(3):273–297. doi: 10.1007/BF00994018.

Cuypers S, Nascetti A, Vergauwen M. 2023. Land use and land cover mapping with VHR and multi-temporal sentinel-2 imagery. Remote Sens. 15(10):2501. doi: 10.3390/rs15102501.

De Jong SM, Van der Meer FD. 2007. Remote sensing image analysis: including the spatial domain. vol. 5. Springer Science & Business Media.

Demir I, Koperski K, Lindenbaum D, Pang G, Huang J, Basu S, Hughes F, Tuia D, Raskar R. 2018. Deepglobe 2018: a challenge to parse the earth through satellite images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, Utah, USA. p. 172–181.

Dong S, Zhuang Y, Yang Z, Pang L, Chen H, Long T. 2020. Land cover classification from VHR optical remote sensing images by feature ensemble deep learning network. IEEE Geosci Remote Sensing Lett. 17(8):1396–1400. doi: 10.1109/LGRS.2019.2947022.

Dornik A, Drăguţ L, Urdea P. 2018. Classification of soil types using geographic object-based image analysis and random forests. Pedosphere. 28(6):913–925. doi: 10.1016/S1002-0160(17)60377-1.

Drăguţ L, Csillik O, Eisank C, Tiede D. 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. ISPRS J Photogramm Remote Sens. 88(100):119–127. doi: 10.1016/j.isprsjprs.2013.11.018.

Drăguţ L, Eisank C, Strasser T. 2011. Local variance for multi-scale analysis in geomorphometry. Geomorphology. 130(3–4):162–172. doi: 10.1016/j.geomorph.2011.03.011.

Duarte D, Nex F, Kerle N, Vosselman G. 2018. Multi-resolution feature fusion for image classification of building damages with convolutional neural networks. Remote Sens. 10(10):1636. doi: 10.3390/rs10101636.

Eckert S, Ghebremicael ST, Hurni H, Kohler T. 2017. Identification and classification of structural soil conservation measures based on very high resolution stereo satellite data. J Environ Manage. 193:592–606. doi: 10.1016/j.jenvman.2017.02.061.

Edwards AW, Cavalli-Sforza LL. 1965. A method for cluster analysis. Biometrics. 21(2):362–375.

Ekanayake E, Ekanayake E, Rathnayake A, Vithana S, Herath H, Godaliyadda G, Ekanayake M. 2018. A semi-supervised algorithm to map major vegetation zones using satellite hyperspectral data. In 2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, Netherlands; IEEE. p. 1–5. doi: 10.1109/WHISPERS.2018.8747025.

Ekaputra F, Sabou M, Serral Asensio E, Kiesling E, Biffl S. 2017. Ontology-based data integration in multi-disciplinary engineering environments: a review. Open J InformSyst. 4(1):1–26.

Fan J, Zeng G, Body M, Hacid MS. 2005. Seeded region growing: an extensive and comparative study. Pattern Recog Lett. 26(8):1139–1156. doi: 10.1016/j.patrec.2004.10.010.

Fan R, Feng R, Wang L, Yan J, Zhang X. 2020. Semi-MCNN: a semisupervised multi-CNN ensemble learning method for urban land cover classification using submeter HRRS images. IEEE J Sel Top Appl Earth Observations Remote Sensing. 13:4973–4987. doi: 10.1109/JSTARS.2020.3019410.

Fang B, Kou R, Pan L, Chen P. 2019. Category-sensitive domain adaptation for land cover mapping in aerial scenes. Remote Sens. 11(22):2631. doi: 10.3390/rs11222631.
Fisher P. 1997. The pixel: a snare and a delusion. Int J Remote Sens. 18(3):679–685. doi: 10.1080/014311697219015.
Foody GM, Mathur A. 2004. A relative evaluation of multiclass image classification by support vector machines. IEEE Trans Geosci Remote Sensing. 42(6):1335–1343. doi: 10.1109/TGRS.2004.827257.
Franklin S, Hall R, Moskal L, Maudie A, Lavigne M. 2000. Incorporating texture into classification of forest species composition from airborne multispectral images. Int J Remote Sens. 21(1):61–79. doi: 10.1080/014311600210993.
Friedl MA, Brodley CE, Strahler AH. 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. IEEE Trans Geosci Remote Sensing. 37(2):969–977. doi: 10.1109/36.752215.
Fu G, Liu C, Zhou R, Sun T, Zhang Q. 2017. Classification for high resolution remote sensing imagery using a fully convolutional network. Remote Sens. 9(5):498. doi: 10.3390/rs9050498.
Ghamisi P, Dalla Mura M, Benediktsson JA. 2015. A survey on spectral–spatial classification techniques based on attribute profiles. IEEE Trans Geosci Remote Sens. 53(5):2335–2353. doi: 10.1109/TGRS.2014.2358934.
Ghita O, Whelan PF. 2002. Computational approach for edge linking. J Electron Imag. 11(4):479–485. doi: 10.1117/1.1501574.
Godwin C, Chen G, Singh KK. 2015. The impact of urban residential development patterns on forest carbon density: an integration of lidar, aerial photography and field mensuration. Landscape Urban Plann. 136:97–109. doi: 10.1016/j.landurbplan.2014.12.007.
Gruber TR. 1995. Toward principles for the design of ontologies used for knowledge sharing? Int J Hum Comput Stud. 43(5–6):907–928. doi: 10.1006/ijhc.1995.1081.
Gu H, Han Y, Yang Y, Li H, Liu Z, Soergel U, Blaschke T, Cui S. 2018. An efficient parallel multi-scale segmentation method for remote sensing imagery. Remote Sens. 10(4):590. doi: 10.3390/rs10040590.
Guindon B. 1997. Computer-based aerial image understanding: a review and assessment of its application to planimetric information extraction from very high resolution satellite images. Can J Remote Sens. 23(1):38–47. doi: 10.1080/07038992.1997.10874676.
Guo J, Zhou H, Zhu C. 2013. Cascaded classification of high resolution remote sensing images using multiple contexts. Inf Sci. 221:84–97. doi: 10.1016/j.ins.2012.09.024.
Haklay M, Weber P. 2008. Openstreetmap: user-generated street maps. IEEE Pervasive Comput. 7(4):12–18. doi: 10.1109/MPRV.2008.80.
Hay G, Niemann K, McLean G. 1996. An object-specific image-texture analysis of h-resolution forest imagery. Remote Sens Environ. 55(2):108–122. doi: 10.1016/0034-4257(95)00189-1.
He H, Liang T, Hu D, Yu X. 2016a. Remote sensing clustering analysis based on object-based interval modeling. Comput Geosci. 94:131–139. doi: 10.1016/j.cageo.2016.06.006.
He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA. p. 770–778.
Hegyi A, Vernica MM, Drăguț L. 2020. An object-based approach to support the automatic delineation of magnetic anomalies. Archaeological Prospection. 27(1):3–12. doi: 10.1002/arp.1752.
Hofmann P, Blaschke T, Strobl J. 2011. Quantifying the robustness of fuzzy rule sets in object-based image analysis. Int J Remote Sens. 32(22):7359–7381. doi: 10.1080/01431161.2010.523727.
Homer CH, Fry JA, Barnes CA. 2012. The national land cover database. US Geological Survey Fact Sheet. 3020(4):1–4.
Hossain MD, Chen D. 2019. Segmentation for object-based image analysis (Obia): a review of algorithms and challenges from remote sensing perspective. ISPRS J Photogramm Remote Sens. 150:115–134. doi: 10.1016/j.isprsjprs.2019.02.009.
Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. 2017. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA. p. 4700–4708.
Huang X, Zhang L. 2013. An svm ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. IEEE Trans Geosci Remote Sensing. 51(1):257–272. doi: 10.1109/TGRS.2012.2202912.
Imaging D. 2002. Recognition, version 2.1. Germany: Definiens Imaging GmbH, München.
Jain R, Kasturi R, Schunck BG. 1995. Machine vision. vol. 5. New York: McGraw-Hill.
Jin Q, Xu E, Zhang X. 2022. A fusion method for multisource land cover products based on superpixels and statistical extraction for enhancing resolution and improving accuracy. Remote Sensing. 14(7):1676. doi: 10.3390/rs14071676.

Johnson B, Xie Z. 2011. Unsupervised image segmentation evaluation and refinement using a multi-scale approach. ISPRS J Photogramm Remote Sens. 66(4):473–483. doi: 10.1016/j.isprsjprs.2011.02.006.
Kamal M, Phinn S. 2011. Hyperspectral data for mangrove species mapping: a comparison of pixel-based and object-based approach. Remote Sens. 3(10):2222–2242. doi: 10.3390/rs3102222.
Kettig RL, Landgrebe D. 1976. Classification of multispectral image data by extraction and classification of homogeneous objects. IEEE Trans Geosci Electron. 14(1):19–26. doi: 10.1109/TGE.1976.294460.
Khatami R, Mountrakis G, Stehman SV. 2016. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. Remote Sens Environ. 177:89–100. doi: 10.1016/j.rse.2016.02.028.
Kiryati N, Eldar Y, Bruckstein AM. 1991. A probabilistic hough transform. Pattern Recognit. 24(4):303–316. doi: 10.1016/0031-3203(91)90073-E.
Krizhevsky A, Sutskever I, Hinton GE. 2012. Imagenet classification with deep convolutional neural networks. Adv Neur Inform Process Syst. 25.
Kuhn W, Raubal M, Gärdenfors P. 2007. Cognitive semantics and spatio-temporal ontologies. Spat Cognit Comput. 7(1):3–12. doi: 10.1080/13875860701337835.
Kundu R. 2022. Image processing: techniques, types, and applications; 2023. [accessed 2023 Aug 12]. https://www.v7labs.com/blog/image-processing-guide.
Kwak T, Kim Y. 2023. Semi-supervised land cover classification of remote sensing imagery using cyclegan and efficientnet. KSCE J Civ Eng. 27(4):1760–1773. doi: 10.1007/s12205-023-2285-0.
Kwenda C, Gwetu M, Fonou-Dombeu JV. 2023. Ontology with deep learning for forest image classification. Applied Sciences. 13(8):5060. doi: 10.3390/app13085060.
Laliberte AS, Rango A. 2009. Texture and scale in object-based analysis of subdecimeter resolution unmanned aerial vehicle (UAV) imagery. IEEE Trans Geosci Remote Sensing. 47(3):761–770. doi: 10.1109/TGRS.2008.2009355.
Lang S. 2008. Object-based image analysis for remote sensing applications: modeling reality–dealing with complexity. Object Based Image Anal. :3–27. Springer.
Lang S, Blaschke T. 2006. Bridging remote sensing and GIS–what are the main supportive pillars. In Proceedings of the 1st International Conference on Object-Based Image Analysis, Salzburg, Austria. p. 4–5.
Lang S, Hay GJ, Baraldi A, Tiede D, Blaschke T. 2019. Geobia achievements and spatial opportunities in the era of big earth observation data. IJGI. 8(11):474. doi: 10.3390/ijgi8110474.
Lang S, Kienberger S, Tiede D, Hagenlocher M, Pernkopf L. 2014. Geons–domain-specific regionalization of space. Cartograph Geograph Inform Sci. 41(3):214–226. doi: 10.1080/15230406.2014.902755.
Lang S, Zeil P, Kienberger S, Tiede D. 2008. Geons–policy-relevant geo-objects for monitoring high-level indicators. In: Car A, Griesebner G, Strobl J. editors. Geospatial Crossroads@ GI_Forum. Proceedings of the Second Geoinformatics Forum. Salzburg, Heidelberg: Wichmannn.
Larochelle H. 2020. Few-shot learning. Computer vision: a reference guide. p. 1–4. Springer.
Lasaponara R, Leucci G, Masini N, Persico R, Scardozzi G. 2016. Towards an operative use of remote sensing for exploring the past using satellite data: the case study of Hierapolis (Turkey). Remote Sens Environ. 174:148–164. doi: 10.1016/j.rse.2015.12.016.
Lei T, Li L, Lv Z, Zhu M, Du X, Nandi AK. 2021. Multi-modality and multi-scale attention fusion network for land cover classification from VHR remote sensing images. Remote Sensing. 13(18):3771. doi: 10.3390/rs13183771.
Li A, Lu Z, Wang L, Xiang T, Wen JR. 2017. Zero-shot scene classification for high spatial resolution remote sensing images. IEEE Trans Geosci Remote Sensing. 55(7):4157–4167. doi: 10.1109/TGRS.2017.2689071.
Li Y, Ouyang S, Zhang Y. 2022. Combining deep learning and ontology reasoning for remote sensing image semantic segmentation. Knowledge Based Syst. 243:108469. doi: 10.1016/j.knosys.2022.108469.
Li Z, Snavely N. 2018. Megadepth: learning single-view depth prediction from internet photos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, USA. p. 2041–2050.
Littlestone N. 1988. Learning quickly when irrelevant attributes abound: a new linear-threshold algorithm. Mach Learn. 2(4):285–318. doi: 10.1007/BF00116827.
Liu D, Kelly M, Gong P. 2006. A spatial–temporal approach to monitoring forest disease spread using multi-temporal high spatial resolution imagery. Remote Sens Environ. 101(2):167–180. doi: 10.1016/j.rse.2005.12.012.
Lu D, Li G, Moran E, Hetrick S. 2013. Spatiotemporal analysis of land-use and land-cover change in the brazilian amazon. Int J Remote Sens. 34(16):5953–5978. doi: 10.1080/01431161.2013.802825.

Lucchese L, Mitra SK. 2001. Colour image segmentation: a state-of-the-art survey. Proc Indian Natl Sci Acad Part A. 67(2):207–222.
Luhmann T, Robson S, Kyle S, Boehm J. 2019. Close-range photogrammetry and 3d imaging. In: Luhmann T, Robson S, Kyle S, Boehm J, editors. Close-range photogrammetry and 3D imaging. Berlin, Germany: de Gruyter.
Martin DR, Fowlkes CC, Malik J. 2004. Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Trans Pattern Anal Mach Intell. 26(5):530–549. doi: 10.1109/TPAMI.2004.1273918.
Merciol F, Faucqueur L, Damodaran BB, Rémy PY, Desclée B, Dazin F, Lefèvre S, Sannier C. 2018. Geobia at the terapixel scale: from VHR satellite images to small woody features at the pan-European level. In: GEOBIA 2018-From Pixels to Ecosystems and Global Sustainability; Montpellier, France.
Ming D, Li J, Wang J, Zhang M. 2015. Scale parameter selection by spatial statistics for geobia: using mean-shift based multi-scale segmentation as an example. ISPRS J Photogramm Remote Sens. 106:28–41. doi: 10.1016/j.isprsjprs.2015.04.010.
Mirghasemi S, Rayudu R, Zhang M. 2013. A new image segmentation algorithm based on modified seeded region growing and particle swarm optimization. In 2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013), Wellington, New Zealand. IEEE. p. 382–387.
Mowrer HT, Congalton RG. 2000. Quantifying spatial uncertainty in natural resources: theory and applications for GIS and remote sensing. Boca Raton, FL: CRC Press.
Nock R, Nielsen F. 2004. Statistical region merging. IEEE Trans Pattern Anal Mach Intell. 26(11):1452–1458. doi: 10.1109/TPAMI.2004.110.
Noy NF, McGuinness DL. 2001. Ontology development 101: a guide to creating your first ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.
Osio A, Lefèvre S, Ogao P, Ayugi S. 2018. Obia-based monitoring of riparian vegetation applied to the identification of degraded acacia Xanthophloea along Lake Nakuru, Kenya. In: GEOBIA 2018-From Pixels to Ecosystems and Global Sustainability, Montpellier, France. p. 18–22.
Pal M, Mather PM. 2003. An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sens Environ. 86(4):554–565. doi: 10.1016/S0034-4257(03)00132-9.
Pan J, Hu X, Li P, Li H, He W, Zhang Y, Lin Y. 2016. Domain adaptation via multi-layer transfer learning. Neurocomputing. 190:10–24. doi: 10.1016/j.neucom.2015.12.097.
Panawong J, Kaewboonma N, Chansanam W. 2018. Building an ontology of flora of Thailand for developing semantic electronic dictionary. In AIP Conference Proceedings; Maharashtra, India: AIP Publishing LLC. p. 020118, vol. 2016.
Pati C, Panda AK, Tripathy AK, Pradhan SK, Patnaik S. 2020. A novel hybrid machine learning approach for change detection in remote sensing images. Eng Sci Technol Int J. 23(5):973–981. doi: 10.1016/j.jestch.2020.01.002.
Pekkarinen A. 2002. Image segment-based spectral features in the estimation of timber volume. Remote Sens Environ. 82(2–3):349–359. doi: 10.1016/S0034-4257(02)00052-4.
Peng H, Long F, Ding C. 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell. 27(8):1226–1238. doi: 10.1109/TPAMI.2005.159.
Pipaud I, Lehmkuhl F. 2017. Object-based delineation and classification of alluvial fans by application of mean-shift segmentation and support vector machines. Geomorphology. 293:178–200. doi: 10.1016/j.geomorph.2017.05.013.
Powers RP, Hermosilla T, Coops NC, Chen G. 2015. Remote sensing and object-based techniques for mapping fine-scale industrial disturbances. Int J Appl Earth Obs Geoinf. 34:51–57. doi: 10.1016/j.jag.2014.06.015.
Qin R. 2015. A mean shift vector-based shape feature for classification of high spatial resolution remotely sensed imagery. IEEE J Sel Top Appl Earth Observations Remote Sensing. 8(5):1974–1985. doi: 10.1109/JSTARS.2014.2357832.
Qin R, Liu T. 2022. A review of landcover classification with very-high resolution remotely sensed optical images—analysis unit, model scalability and transferability. Remote Sensing. 14(3):646. doi: 10.3390/rs14030646.
Radoux J, Defourny P. 2008. Quality assessment of segmentation results devoted to object-based classification. In: Blaschke T, Lang S, Hay GJ, editors. Object-based image analysis. Berlin, Heidelberg: Springer; p. 257–271.

Redmon J, Divvala S, Girshick R, Farhadi A. 2016. You only look once: unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA. p. 779–788.

Sahin K, Ulusoy I. 2013. Automatic multi-scale segmentation of high spatial resolution satellite images using watersheds. In 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS, Melbourne, Australia; IEEE. p. 2505–2508.

Sarker IH. 2021. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci. 2(6):420. doi: 10.1007/s42979-021-00815-1.

Schäfer E, Heiskanen J, Heikinheimo V, Pellikka P. 2016. Mapping tree species diversity of a tropical montane forest by unsupervised clustering of airborne imaging spectroscopy data. Ecol Indic. 64:49–58. doi: 10.1016/j.ecolind.2015.12.026.

Schmitt M, Ahmadi SA, Hänsch R. 2021. There is no data like more data-current status of machine learning datasets in remote sensing. In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium; IEEE. p. 1206–1209.

Sertel E, Ekim B, Ettehadi Osgouei P, Kabadayi ME. 2022. Land use and land cover mapping using deep learning based segmentation approaches and VHR worldview-3 images. Remote Sen. 14(18):4558. doi: 10.3390/rs14184558.

Smith G, Hazeu G. 2015. Review and follow-up alignment of technical outputs (report task 5-3). assistance to the EEA in the production of the new corine land cover (CLC) inventory including the support to the harmonisation of national monitoring for integration at pan-European level. EEA-European Environment Agency. Report No:.

Souza-Filho PWM, Nascimento WR, Jr., Santos DC, Weber EJ, Silva Jr RO, Siqueira JO. 2018. A geobia approach for multitemporal land-cover and land-use change analysis in a tropical watershed in the southeastern Amazon. Remote Sens. 10(11):1683. doi: 10.3390/rs10111683.

Su T. 2017. A novel region-merging approach guided by priority for high resolution image segmentation. Remote Sens Lett. 8(8):771–780. doi: 10.1080/2150704X.2017.1320441.

Sui L, Kang J, Yang X, Wang Z, Wang J. 2020. Inconsistency distribution patterns of different remote sensing land-cover data from the perspective of ecological zoning. Open Geosci. 12(1):324–341. doi: 10.1515/geo-2020-0014.

Sun F, Fang F, Wang R, Wan B, Guo Q, Li H, Wu X. 2020. An impartial semi-supervised learning strategy for imbalanced classification on VHR images. Sensors. 20(22):6699. doi: 10.3390/s20226699.

Szegedy C, Ioffe S, Vanhoucke V, Alemi A. 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, California, USA; vol. 31. doi: 10.1609/aaai.v31i1.11231.

Tehrany MS, Pradhan B, Jebuv MN. 2014. A comparative assessment between object and pixel-based classification approaches for land use/land cover mapping using spot 5 imagery. Geocarto Int. 29(4):351–369. doi: 10.1080/10106049.2013.768300.

Teruggi S, Grilli E, Russo M, Fassi F, Remondino F. 2020. A hierarchical machine learning approach for multi-level and multi-resolution 3d point cloud classification. Remote Sens. 12(16):2598. doi: 10.3390/rs12162598.

Theodoridou I, Karteris M, Mallinis G, Tsiros E, Karteris A. 2017. Assessing the benefits from retrofitting green roofs in Mediterranean, using environmental modelling, GIS and very high spatial resolution remote sensing data: the example of Thessaloniki, Greece. Procedia Environ Sci. 38:530–537. doi: 10.1016/j.proenv.2017.03.117.

Tompoulidou M, Gitas I, Polychronaki A, Mallinis G. 2016. A geobia framework for the implementation of national and international forest definitions using very high spatial resolution optical satellite data. Geocarto Int. 31(3):342–354. doi: 10.1080/10106049.2015.1047470.

Too J, Abdullah AR. 2021. A new and fast rival genetic algorithm for feature selection. J Supercomput. 77(3):2844–2874. doi: 10.1007/s11227-020-03378-9.

Van Beijma S, Comber A, Lamb A. 2014. Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data. Remote Sens Environ. 149:118–129. doi: 10.1016/j.rse.2014.04.010.

Van der Linden S, Hostert P. 2009. The influence of urban structures on impervious surface maps from airborne hyperspectral data. Remote Sens Environ. 113(11):2298–2305. doi: 10.1016/j.rse.2009.06.004.

Vaz DA, Sarmento PT, Barata MT, Fenton LK, Michaels TI. 2015. Object-based dune analysis: automated dune mapping and pattern characterization for Ganges Chasma and gale crater, mars. Geomorphology. 250:128–139. doi: 10.1016/j.geomorph.2015.08.021.

Venkataramanan A, Laviale M, Figus C, Usseglio-Polatera P, Pradalier C. 2021. Tackling inter-class similarity and intra-class variance for microscopic image-based classification. In International Conference on Computer Vision Systems; Springer. p. 93–103.

Verma OP, Hanmandlu M, Susan S, Kulkarni M, Jain PK. 2011. A simple single seeded region growing algorithm for color image segmentation using adaptive thresholding. In 2011 International Conference on Communication Systems and Network Technologies, Katra, Jammu and Kashmir, India; IEEE. p. 500–503.

Vermeulen D, van Niekerk A. 2016. Evaluation of a worldview-2 image for soil salinity monitoring in a moderately affected irrigated area. J Appl Remote Sens. 10(2):026025. doi: 10.1117/1.JRS.10.026025.

Vickers NJ. 2017. Animal communication: when I'm calling you, will you answer too? Curr Biol. 27(14): R713–R715. doi: 10.1016/j.cub.2017.05.064.

Vincent L, Soille P. 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Trans Pattern Anal Machine Intell. 13(6):583–598. doi: 10.1109/34.87344.

Vogels MF, de Jong SM, Sterk G, Addink EA. 2017. Agricultural cropland mapping using black-and-white aerial photography, object-based image analysis and random forests. Int J Appl Earth Obs Geoinf. 54: 114–123. doi: 10.1016/j.jag.2016.09.003.

Wang H, Wang Y, Zhang Q, Xiang S, Pan C. 2017. Gated convolutional neural network for semantic segmentation in high-resolution images. Remote Sens. 9(5):446. doi: 10.3390/rs9050446.

Wang M, Li R. 2014. Segmentation of high spatial resolution remote sensing imagery based on hard-boundary constraint and two-stage merging. IEEE Trans Geosci Remote Sens. 52(9):5712–5725.

Wang Z, Jensen JR, Im J. 2010. An automatic region-based image segmentation algorithm for remote sensing applications. Environment Model Software. 25(10):1149–1165. doi: 10.1016/j.envsoft.2010.03.019.

Wang Z, Nasrabadi NM, Huang TS. 2015. Semisupervised hyperspectral classification using task-driven dictionary learning with Laplacian regularization. IEEE Trans Geosci Remote Sensing. 53(3):1161–1173. doi: 10.1109/TGRS.2014.2335177.

Webb AR. 2003. Statistical pattern recognition. Hoboken, NJ: John Wiley & Sons.

Wu L, Wang Y, Long J, Liu Z. 2015. A non-seed-based region growing algorithm for high resolution remote sensing image segmentation. In International Conference on Image and Graphics, Tianjin, China; Springer. p. 263–277.

Wu Y, Zhang Z, Wang G. 2019. Unsupervised deep feature transfer for low resolution image classification. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, South Korea. p. 0–0.

Xiang S, Nie F, Zhang C. 2008. Learning a Mahalanobis distance metric for data clustering and classification. Pattern Recognit. 41(12):3600–3612. doi: 10.1016/j.patcog.2008.05.018.

Xie Z, Roberts C, Johnson B. 2008. Object-based target search using remotely sensed data: a case study in detecting invasive exotic Australian pine in south Florida. ISPRS J Photogramm Remote Sens. 63(6): 647–660. doi: 10.1016/j.isprsjprs.2008.04.003.

Yin X, Yang W, Xia GS, Dong L. 2014. Semi-supervised feature learning for remote sensing image classification. In 2014 IEEE Geoscience and Remote Sensing Symposium, Quebec City, Canada; IEEE. p. 1261–1264.

Yosinski J, Clune J, Bengio Y, Lipson H. 2014. How transferable are features in deep neural networks? Adv Neur Inform Process Syst. :27.

Yu Q, Gong P, Clinton N, Biging G, Kelly M, Schirokauer D. 2006. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. Photogramm Eng Remote Sensing. 72(7):799–811. doi: 10.14358/PERS.72.7.799.

Yuan X, Shi J, Gu L. 2021. A review of deep learning methods for semantic segmentation of remote sensing imagery. Expert Syst Appl. 169:114417. doi: 10.1016/j.eswa.2020.114417.

Zhang AZ, Sun GY, Liu SH, Wang ZJ, Wang P, Ma JS. 2017. Multi-scale segmentation of very high resolution remote sensing image based on gravitational field and optimized region merging. Multimed Tools Appl. 76(13):15105–15122. doi: 10.1007/s11042-017-4558-4.

Zhang C, Xie Z, Selch D. 2013. Fusing Lidar and digital aerial photography for object-based forest mapping in the Florida everglades. GIScience Remote Sens. 50(5):562–573. doi: 10.1080/15481603.2013.836807.

Zhang F, Yang X. 2020. Improving land cover classification in an urbanized coastal area by random forests: the role of variable selection. Remote Sens Environ. 251:112105. doi: 10.1016/j.rse.2020.112105.

Zhang L, Zhang L, Du B. 2016. Deep learning for remote sensing data: a technical tutorial on the state of the art. IEEE Geosci Remote Sens Mag. 4(2):22–40. doi: 10.1109/MGRS.2016.2540798.

Zhang X, Han L, Han L, Zhu L. 2020. How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery? Remote Sens. 12(3):417. doi: 10.3390/rs12030417.

Zhang X, Zhou X, Lin M, Sun J. 2018. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, USA. p. 6848–6856.

Zhang Z, Luo C, Wu H, Chen Y, Wang N, Song C. 2022. From individual to whole: reducing intra-class variance by feature aggregation. Int J Comput Vis. 130(3):800–819. doi: 10.1007/s11263-021-01569-2.

Zhou G, Xu J, Chen W, Li X, Li J, Wang L. 2023. Deep feature enhancement method for land cover with irregular and sparse spatial distribution features: a case study on open-pit mining. IEEE Trans Geosci Remote Sens. 61:1–20. doi: 10.1109/TGRS.2023.3241331.

Zhu L, Jin G, Gao D. 2021. Integrating land-cover products based on ontologies and local accuracy. Information. 12(6):236. doi: 10.3390/info12060236.

2.2.2 Conclusion

The study conducted a comprehensive analysis of Geographic Object-Based Image Analysis (GEOBIA) techniques employed in the identification and classification of forest images obtained from satellite imagery. This study addresses several challenges encountered in the field of Geographic Object-Based Image Analysis (GEOBIA). These challenges include difficulties in effectively addressing segmentation scales, the subjective nature of the analysis, and limitations in managing large-scale geospatial data. The research also outlines the potential future possibilities that studies in GEOBIA (Geographic Object-Based Image Analysis) might pursue. Additionally, it proposed a cutting-edge ontological framework for the classification of satellite forest images. The study suggests that using XAI (Explainable Artificial Intelligence) with ontologies in the classification of forest photos can enhance expert confidence in the outcomes. The recommended approach provides a detailed explanation of the steps involved in reaching a conclusion, in contrast to the opaque nature of deep neural network models.

3 Satellite Forest Image Segmentation

Image segmentation is a key process for subsequent image classification. This chapter presents published works in the segmentation of satellite forest images.

3.1 Hybrid Learning Model for Satellite Forest Image Segmentation

3.1.1 Introduction

This section introduces a research paper that presented the segmentation of satellite forest images into regions of forest and non-forest areas. The work employed ResNet50 deep learning technique to generate a set of features for the Random Forest algorithm to perform the segmentation process. The model was subjected to a fine-tuning process to increase its efficiency. The proposed strategy in the study contributed to better performance in the accuracy of the segmentation results.

Hybrid Learning Model for Satellite Forest Image Segmentation

Clopas Kwenda( ), Mandlenkosi Victor Gwetu, and Jean Vincent Fonou-Dombeu

School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa
221072651@stu.ukzn.ac.za, {gwetum,fonoudombeuj}@ukzn.ac.za

Abstract. Image segmentation is an essential image processing technique as the quality of individual object detection significantly affects subsequent global image classification accuracy. The segmentation process can be performed by a varying number of different algorithms, but to date, these different algorithms are not yet able to guarantee a level of performance similar to or superior to human capability. This study adopts a supervised approach toward satellite forest image segmentation. The proposed model used a feature vector obtained through transfer learning from ResNet50; these features were then passed to a Random Forest for segmentation. The satellite images used for training and testing were obtained from the Land Cover Classification Truck in DeepGlobe Challenge. Metrics such as precision, recall, F1-Score, accuracy, Root Mean Square Error (RMSE), and Mean Average Error (MAE) were used to assess the performance of the model. The model achieved a testing accuracy of 94%, RMSE value of 0.2499, and MAE value of 5.92. For detecting forest areas the proposed model obtained a precision of 0.94, recall of 0.96, and F1-Score of 0.95. For non-forest areas, the proposed model achieved a precision of 0.93, recall of 0.89, and F1-Score of 0.91.

Keywords: Segmentation · Supervised approach · ResNet50 · Random Forest · Remote Sensing Image

1 Introduction

Image segmentation is a process whereby an image is partitioned into separate regions, which ideally relate to different real-world objects [1]. Such a process is of paramount importance for subsequent computational image content analysis and understanding. Hence, image segmentation quality significantly affects subsequent image classification accuracy. The segmentation process can be performed by a varying number of different algorithms, but to date, a satisfactory level of high performance by these different algorithms has not yet been realized. This can be attributed to different standards of what constitutes a good segmentation, the sheer complexity of the segmentation context as well as a possible mismatch between the applied segmentation technique and its domain use

L. Rutkowski et al. (Eds.): ICAISC 2023, LNAI 14126, pp. 37–47, 2023.
https://doi.org/10.1007/978-3-031-42508-0_4

case. Given the complexity and application dependence of the image segmentation task, this study advocates for the design of a segmentation technique that leverages algorithm variance. Segmentation quality is assessed by measuring the discrepancy between the delineated image region (DIR) and the actual image region (AIR) of the scene image.

Unsupervised methods (empirical goodness methods) and Supervised methods (empirical discrepancy methods) [2] are the dominant evaluation algorithms used in remote sensing images. Supervised methods evaluate a segmented image object with a reference to a gold standard image (i.e a manually segmented image). The quality of the segmented image is determined by the similarity between gold standard image and the segmented image. The main advantage of this approach is that there is a direct comparison between the manually generated image and the segmented image. Unsupervised methods depend solely on the segmented image, i.e. it does not require to be compared with manually segmented reference image [1]. The main advantage of Unsupervised evaluation assessment methods is that they do not require to be assessed against a truth value (i.e. manually segmented reference image). Therefore they are most suited for general-purpose segmentation applications. However, Cheng [3] pointed out that, whenever a sound ground truth is established, supervised methods are desirable for segmentation assessment evaluation. It is against this backdrop that this study has adopted a supervised approach toward satellite forest image segmentation. The proposed model used a feature vector obtained through transfer learning from ResNet50; these features were then passed to a Random Forest for segmentation. The satellite images used for training and testing were obtained from the Land Cover Classification Truck in DeepGlobe Challenge. Metrics such as precision, recall, F1-Score, accuracy, Root Mean Square Error (RMSE), and Mean Average Error (MAE) were used to assess the performance of the model. The model achieved a testing accuracy of 94%, RMSE value of 0.2499, and MAE value of 5.92. For detecting forest areas the proposed model obtained a precision of 0.94, recall of 0.96, and F1-Score of 0.95. For non-forest areas, the proposed model achieved a precision of 0.93, recall of 0.89, and F1-Score of 0.91.

The rest of the paper is structured as follows. Section 2 discusses related work. The materials and methods used in the study are presented in Sect. 3. Section 4 describes and discusses the results of the study. The paper is concluded in Sect. 5.

2 Related Work

A study [4] investigated the use of auto-encoders to estimate forest biomass on Landsat8 and LiDAR data sets. The auto-encoders outperformed the traditional machine learning algorithms such as the k-nearest neighbor, support vector regression, and multiple step-wise linear regression by 1% to 7% in terms of relative RMSE. The main limitation of their study was that there was no mechanism or method for selecting optimal predictor variables, and separate estimates for different forest types, hence the study obtained lower RMSE values. InceptionV3 and GoogleNet architectures were successfully implemented to

estimate forest-above biomass from LiDAR data with RMSE of 26% and bias of 0.7% respectively [5]. The study demonstrated that the use of deep learning methods such as CNN for interpreting LiDAR datasets is an improvement upon traditional methods for area-based predictions of forest attributes. However such improvements brought about some drawbacks. The CNN deep learning model requires large amounts of training data, effort, and time to perform the modeling process. As a result, it is upon's modeler's judgment to decide whether the improvements in the model's performance are worth the effort required to train the model. The proposed model in this study can successfully produce good results on limited data set as it is only the RF section that performs the key segmentation process and it performs best on the limited dataset.

In another study [6], the authors came up with a multi-task recurrent CNN that integrated data from several sources, including aerial and satellite image time series, and climate data for classifying different forest cover types and forest proprieties such as above-ground biomass, canopy cover, basal area and quadratic mean diameter. The multi-task method outclassed support vector machines and random forests. Since the model was purely based on CNN, the model required a large amount of training data, and as result, the model failed to make predictions of hardwood forests of the Eastern US due to a lack of images for specific forest types. The model proposed in this study can produce good results on limited datasets.

A study [7] implemented a deep convolutional encoder-decoder Segnet model to distinguish between crops, weeds, and background. A line detection algorithm was used to determine the row of the crops, then the distance of both the crops and weeds from the detected line was fed into a random forest algorithm to label the plant as either weed or crop. The major limitation of this study was the weak features adopted for crop detection which resulted in the misidentification of crops and weeds. To address this issue, a hybrid model based on CNN was used, which excels at producing features required for image segmentation and classification [8].

A U-net deep CNN was used [9] for orchard tree segmentation using aerial images. The aim was to detect and localize a canopy of orchard trees under various conditions. The study en-counted a lot of false positives based on the segmentation results. This could be attributed to high compression of training images which resulted in the loss of vital information. The proposed model includes an image pre-processing phase in which all input images are resized to 512×512, which is large enough for the CNN model to capture all vital features [10].

A model based on semantic segmentation and lidar odometry for the tree diameter estimation was proposed [11]. Virtual reality tool was used to label 3D scans that were in turn used to train a semantic segmentation network. The resulting masks were used to compute a trellis graph that uniquely identifies each instance and extracts relevant features for the SLAM module. The model was able to automatically generate tree diameter estimations. The study encountered the challenge of quantifying the performance of the traditional SLAM algorithm

as it was difficult to obtain ground truth measurements of trajectories in the natural environment. To address this issue, ground truth measurements of easily obtained landmark shapes can be used to benchmark various algorithms. The proposed model made use of easily accessible image labels from [12].

3 Materials and Methods

This section presents the ResNet50 architecture used in this study for feature extraction, the RF method employed for the segmentation of images, the metrics for evaluating segmentation results, and describes various operations such as image processing and feature extraction, fine-tuning, sorting of features by importance and performance test of the model.

3.1 Random Forest

Random forest (RF) is regarded as an ensemble classification method that uses a set of classifiers instead of one classifier to classify a new set of data points by considering their vote predictions. Bagging and Boosting algorithms are also under the category of ensemble classification techniques. Boosting uses the concept of iterative retraining to update the weights of incorrectly classified samples. However, it is sensitive to noise, very slow, and can over-train [13]. The Bagging algorithm draws many bootstrap samples from training data and for each bootstrap, a tree is constructed. The idea is that the successive trees are constructed independently from earlier trees and a simple majority vote is taken for prediction [14]. RF is an advanced version of Bagging which is simply described as the collection of tree-structured classifiers [15]. RF splits each node by considering the absolute best split from a randomly chosen subset of predictors at that particular node. A new training data set referred to as the bootstrap is created from the original data set. The random feature selection is then used to grow a tree. RF is initialized by setting two parameters that are of paramount importance, that is, the number of trees to grow, referred to as N, and the number of variables m to split at each node. Then the N bootstrap is obtained by taking two-thirds of the training data, whilst the remaining one-third of the training data that is referred to as out-of-bag (OOB) data is reserved for testing. An unpruned tree from each bootstrap is constructed with a constraint that, each node is based on what is considered the best split (GINI Index) from randomly chosen subset predictor variables. The GINI index also referred to as GINI impurity measures class homogeneity and is expressed as follows.

$$G = \sum_{m=1}^{J} c_m(1 - c_m) \tag{1}$$

where J is the number of classes and c_m corresponds to a set of items labelled with classes m $\in$ 1, 2, ..., J.

3.2 ResNet50

The ResNet50 model was formulated to provide solutions to difficulties in the training process of deep CNN and to reduce saturation and degradation problems. ResNet50 architecture has been developed to overcome the degradation problem by using residual learning. It is an extremely deep type of CNN with 48 convolutional layers, one Maxpooling layer, and one Average pooling layer. An input instance and an output instance are summed up such that the original mapping function

$$H(x) = F(x) - x \tag{2}$$

is redefined as:

$$H(x) = F(x) + x \tag{3}$$

The refinement of the mapping function greatly approximates the desired functions while also making learning simple. This reformulation was initiated to mitigate the degradation problem. The redefined mapping function in Eq. 3 is implemented by having feed-forward neural networks with short connections. The short or skip connections are direct connections that skip some of the layers of the model. The shortcut connections carry out identity mapping operations, and the results are added to the outputs of the stacked layers. If the additional layers can be built as identity mappings, the training error of a deeper model should be not greater than that of its shallower counterpart.

3.3 Metrics for Segmentation Evaluation

The authors in [16] revealed that the segmentation of remote sensing images hugely suffers from over-segmentation and under-segmentation problems and there is no standard way of assessing remote sensing segmentation quality. Metrics that are commonly used for evaluation involve accuracy, precision, recall, F1-Score, support, confusion matrix, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) [17].

3.4 Image Preprocessing and Features Extraction

Image pre-processing tasks enhance the quality of image datasets prior to the image segmentation and classification tasks. Generally, this pre-processing task includes image scaling, rotation, and image translation. For this study, all images were resized to 512 × 512 pixels. Due to their architecture, deep learning models generally require that all images in a dataset should be of the same size. CNNs are commonly used for feature extraction. The ResNet50 model which was optimized using transfer learning produced a feature vector with 64 features from the pre-processed original image. Figure 1 shows the distribution of features obtained.

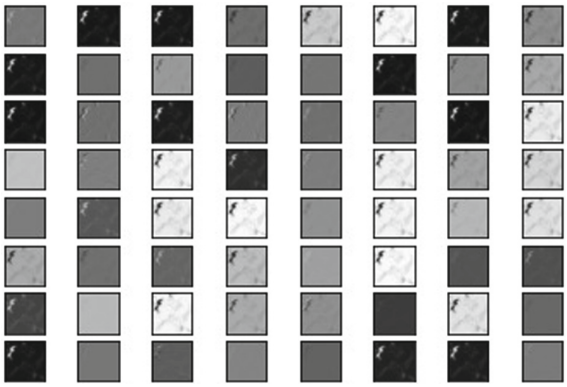


Fig. 1. ResNet50 feature vector

3.5 Fine Tuning

Fine-tuning involves making some modifications to a function or a model in order to improve its effectiveness. The first three layers of ResNet50, up to batch normalization, were considered for feature generation. The layers are shown in Table 1. The batch normalization layer was chosen for tapping the output because the size of images would not have been significantly reduced. The number of estimators in the RF was set to 100. Training image masks were resized to 256×256 to correspond to the output shape of the batch normalization layer.

Table 1. Three ResNet50 layers considered for feature generation

Layer(type)	Output Shape	Parameter
input_1(input layer)	[(None,512,512,3)]	0
conv1_pad(ZeroPadding2D)	(None,518,518,3)	0
conv1_conv(con2D)	(None,256,256,64)	9472
conv1_bn(BathNormalisation)	(None,256,256,64)	256

3.6 Sorting Features by Importance

The stage that was of paramount importance was to identify the features that are more critical than others. Feature importance was used to fish out significant features required for forest image segmentation. The Gini index provides the basis for determining feature importance values. Feature importance in RF is expressed as the decrease in node impurity weighted by the probability likelihood of reaching that node. The probability of reaching a node is determined by the number of samples that vote for the node, divided by the number of samples. The higher the value obtained, the higher the importance of the feature. The study adopted the Scikit-learn Python package for the calculation of the gini indexes. All features with values above 0.00, were considered for the image segmentation process.

3.7 Performance Tests

In order to determine the performance of the proposed model, several tests were conducted. The dataset was split into 80% for training purposes and 20% for testing purposes. Other experiments were performed by replacing ResNet50 with other popular pre-trained models such as InceptionV3, VGG16, and Exception.

4 Experimental Results and Discussions

The dataset, the platform used to carry out this study as well as the experimental results achieved are presented and discussed in this section.

4.1 Platform and Dataset

The experiment was carried out on the Google Colab platform which offers free TPU and GPU resources on the cloud. The GPU acceleration of NVIDIA Tesla was used due to the high computational nature of the experiment. A standard, publicly accessible dataset was obtained from Land Cover Classification Truck in DeepGlobe Challenge [12]. Figure 2 shows an original image and its corresponding labeled mask from the dataset, as used in the study.

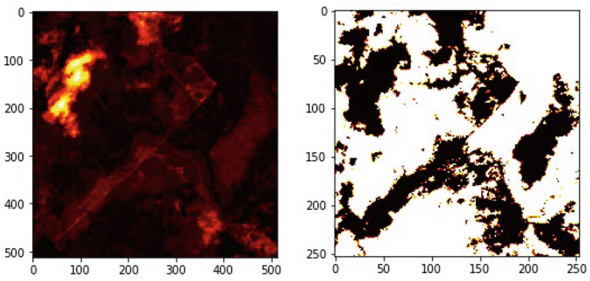


Fig. 2. Extracted RGB-patches and their corresponding masks.

4.2 Results and Discussions

In the experiment of the proposed model, the original image was resized to 512 × 512. Under transfer learning, pre-trained weights of ResNet50 were obtained. ResNet50 architecture produced 64 features when applied to aerial satellite input images. The elements of the 64 feature vectors were taken in as independent variables by the RF classifier to segment forest region areas from non-forest region areas. Several experiments were done by changing the number of classifiers in the RF. The model was then compared against VGG16, Inception, and Xception. Metrics used to evaluate the proposed model were confusion matrix, precision, recall, F1-Score, MAE, and RMSE. The confusion matrix is a two-dimensional table that reflects the performance of a classification algorithm. It is important

Table 2. Metrics for forest areas

Model	Precision	Recall	F1-Score
InceptionV3	0.93	0.94	0.94
Xception	0.93	0.94	0.94
VGG16	0.94	0.97	0.95
Resnet50	**0.94**	**0.96**	**0.95**

for visualizing and summarizing the performance of a classification algorithm. The confusion matrix obtained is presented in Fig. 3. Comparing the segmentation results for each model in Table 2 and Table 3, all the models could effectively segment forest region areas from non-forest region areas, though there are marginal differences in terms of accuracy as presented in Table 4. Both RMSE and MAE are regularly employed for model evaluation studies. [18] argued that RMSE is not a good metric to measure the performance of a model as it gives a misleading average error, hence MAE would be a better metric for such purpose. However, [17] presented that RMSE is not ambiguous in its meaning and hence it is more appropriate to use than MAE when a model's error follows a normal distribution. Also, RMSE satisfies the triangle inequality required for a distance function metric. Because of the argument by [17] the study gives more preference to the RMSE metric.

Table 3. Metrics for non-forest areas

Model	Precision	Recall	F1-Score
InceptionV3	0.89	0.88	0.88
Xception	0.89	0.88	0.88
VGG16	0.93	0.88	0.90
Resnet50	**0.93**	**0.89**	**0.91**

Table 4. Metrics for determining accuracy

Model	Accuracy	RMSE	MAE
InceptionV3	0.9184	0.2857	9.72
Xception	0.9163	0.2893	10.16
VGG16	0.9332	0.2579	5.75
ResNet50	**0.9375**	**0.2499**	**5.92**

The Precision, Recall, and F1-Score values for forest areas (Table 2) are higher than those of non-forest areas (Table 3). InceptionV3 and Xception val-

ues follow each other, hence it can be concluded that these models have got the same computational power for the segmentation task in the context of remote sensing-related images. The hybrid model used by this study achieved a high F1-score (Table 3) as compared to the research done by [19] which achieved an F1-score value of 0.34. Their study was centered on Unet deep learning model for segmenting forest images. The confusion matrix results in Fig. 3 show on the main diagonal that 55 297 pixels were classified correctly while only 3686 pixels were misclassified. In terms of accuracy, Table 4 shows that the hybrid model with ResNet50 produced the best results in performing the segmentation task as it produced an accuracy of 94% and an optimal RMSE value of 0.25. The hybrid model with VGG16 produced closely related results to ResNet50 as it obtained an accuracy of 93% and RMSE of 0.26. For this study VGG16 and ResNet50 had more refined segmentation results, hence they are better at segmenting forest region areas from non-forest region areas. The proposed model achieved an accuracy of 94%, RMSE of 0.2499, and MAE of 5.92 (Table 4). The final segmentation is shown in Fig. 4, where the first, second, and third images are the source, ground truth, and algorithm prediction images, respectively.

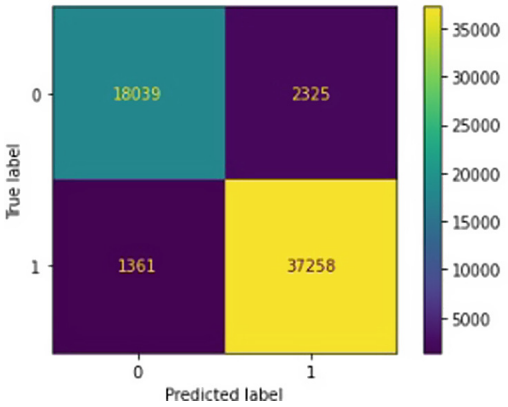


Fig. 3. Confusion Matrix

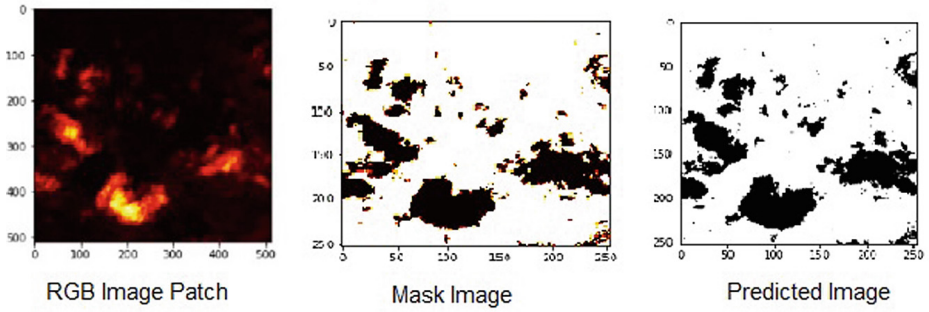


Fig. 4. Forest Image segmentation.

5 Conclusion

In this paper, a hybrid of CNN (ResNet50) and a traditional learning method (Random Forest) was constructed to identify forest areas and non-forest areas. The CNN was employed to produce features which in turn were used by a Random Forest to segment aerial satellite images. The model performance was assessed against other transfer learning models. The proposed model achieved an overall accuracy of 94%. In conclusion, there is no absolute algorithm that is guaranteed to be good in segmentation as it depends on the specific application. The supervised evaluation approach is suitable only if the golden standard truth image is established. The main advantage of this approach is that it produces accurate and reliable results. The fact that unsupervised evaluation approaches do not require standard images and are a low-level data-driven evaluation method, is a factor in the difficulty of obtaining high accuracy and lack of flexibility to accommodate versatility on image features. Future research should consider a model that incorporates ensembling deep learning approaches for feature generation, with an aim of increasing classification accuracy.

References

1. Zhang, H., Fritts, J.E., Goldman, S.A.: Image segmentation evaluation: a survey of unsupervised methods. Comput. Vis. Image Underst. **110**(2), 260–280 (2008)
2. Zhang, Y.J., et al.: A survey on evaluation methods for image segmentation. Pattern Recogn. **29**(8), 1335–1346 (1996)
3. Cheng, J., Bo, Y., Zhu, Y., Ji, X.: A novel method for assessing the segmentation quality of high-spatial resolution remote-sensing images. Int. J. Remote Sens. **35**(10), 3816–3839 (2014)
4. Zhang, L., Shao, Z., Liu, J., Cheng, Q.: Deep learning based retrieval of forest aboveground biomass from combined LiDAR and Landsat 8 data. Remote Sens. **11**(12), 1459 (2019)
5. Ayrey, E., Hayes, D.J.: The use of three-dimensional convolutional neural networks to interpret LiDAR for forest inventory. Remote Sens. **10**(4), 649 (2018)
6. Chang, T., Rasmussen, B.P., Dickson, B.G., Zachmann, L.J.: Chimera: a multi-task recurrent convolutional neural network for forest classification and structural estimation. Remote Sens. **11**(7), 768 (2019)
7. Sa, I., et al.: WeedNet: dense semantic weed classification using multispectral images and MAV for smart farming. IEEE Robot. Autom. Lett. **3**(1), 588–595 (2017)
8. Wang, P., Fan, E., Wang, P.: Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. Pattern Recogn. Lett. **141**, 61–67 (2021)
9. Anagnostis, A., et al.: Orchard mapping with deep learning semantic segmentation. Sensors **21**(11), 3813 (2021)
10. Luke, J.J., Joseph, R., Balaji, M.: Impact of image size on accuracy and generalization of convolutional neural networks. Int. J. Res. Anal. Rev. **6**, 70–80 (2019)
11. Chen, S.W., et al.: SLOAM: semantic LiDAR odometry and mapping for forest inventory. IEEE Robot. Autom. Lett. **5**(2), 612–619 (2020)
12. Quadeer, S.: Forest aerial images for segmentation

13. Halmy, M.W.A., Gessler, P.E.: The application of ensemble techniques for land-cover classification in arid lands. Int. J. Remote Sens. **36**(22), 5613–5636 (2015)
14. Liaw, A., Wiener, M., et al.: Classification and regression by randomForest. R News **2**(3), 18–22 (2002)
15. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
16. Wang, Z., Wang, E., Zhu, Y.: Image segmentation evaluation: a survey of methods. Artif. Intell. Rev. **53**(8), 5637–5674 (2020)
17. Chai, T., Draxler, R.R.: Root mean square error (RMSE) or mean absolute error (MAE)? - arguments against avoiding RMSE in the literature. Geosci. Model Dev. **7**(3), 1247–1250 (2014)
18. Willmott, C.J., Matsuura, K., Robeson, S.M.: Ambiguities inherent in sums-of-squares-based error statistics. Atmos. Environ. **43**(3), 749–752 (2009)
19. Khryashchev, V., Pavlov, V., Ostrovskaya, A., Larionov, R.: Forest areas segmentation on aerial images by deep learning. In: 2019 IEEE East-West Design & Test Symposium (EWDTS), pp. 1–5. IEEE (2019)

3.1.2 Conclusion

This paper presented a hybrid model of the ResNet50 model and Random Forest algorithm constructed to distinguish between forest and non-forest areas. The sole objective of the ResNet50 model was to generate a set of features for the RF algorithm to perform the segmentation process. The model performed better when assessed against other deep learning-based models such as InceptionV3, Xception, and VGG16.

3.2 Hybridizing Deep Neural Networks and Machine learning Models for Aerial Satellite Forest Image Segmentation

3.2.1 Introduction

This paper is an extension of the paper presented in section 3.1. In this study, a hybridization of VGG16 and ResNet50 deep neural networks was employed to extract features from aerial satellite images which were subsequently used by the machine learning classifiers such as RF, LSVM, KNN, LDA, and GNB to segment the image into forest and non-forest regions. The harmonization with the deep neural networks is made to address the challenges of machine learning classifiers of lacking the ability to extract features such as the spatial relationship between pixels and texture which results in subpar segmentation results. This paper has been submitted for publication in MDPI (Remote sensing journal).

Article

Hybridizing Deep Neural Networks and Machine learning Models for Aerial Satellite Forest Image Segmentation

Clopas Kwenda*, Mandlenkosi Gwetu † and Jean Vincent Fonou-Dombeu †

School of Mathematics, Statistics and Computer Science, University of KwaZulu Natal, Pietermaritzburg, 3209, South Africa; gwetum@ukzn.ac.za (M.G.);fonoudombeuj@ukzn.ac.za(J.V.FD)
* Correspondence:221072651@stu.ukzn.ac.za; Tel.: +27-0612039734
† These authors contributed equally to this work.

Abstract: Forests play a pivotal role in mitigating climate change as well as contributing to the socio-economic activities of many countries. Therefore, it is of paramount importance to monitor forest cover. Traditional machine learning classifiers for segmenting images lack the ability to extract features such as the spatial relationship between pixels and texture, resulting in subpar segmentation results when used alone. To address this limitation, this study proposed a novel hybrid approach that combines deep neural networks and machine learning algorithms to segment an aerial satellite image into forest and non-forest regions. Aerial satellite forest image features were first extracted by two deep neural network models, namely, VGG16 and ResNet50. The resulting features are subsequently used by five machine learning classifiers including Random Forest (RF), Linear Support Vector Machines (LSVM), k-nearest neighbor (kNN), Linear Discriminant Analysis (LDA), and Gaussian Naive Bayes (GNB) to perform the final segmentation. The aerial satellite forest images were obtained from a deep globe challenge dataset. The performance of the proposed model was evaluated using metrics such as Accuracy, Jaccard score index, and Root Mean Square Error (RMSE). The experimental results revealed that the RF model achieved the best segmentation results with accuracy, Jaccard score, and RMSE of 94%, 0.913 and 0.245, respectively; followed by LSVM with accuracy, Jaccard score and RMSE of 89%, 0.876, 0.332, respectively. The LDA took the third position with accuracy, Jaccard score, and RMSE of 88%, 0.834, and 0.351, respectively, followed by GNB with accuracy, Jaccard score, and RMSE of 88%, 0.837, and 0.353, respectively. The kNN occupied the last position with accuracy, Jaccard score, and RMSE of 83%, 0.790, and 0.408, respectively. The experimental results also revealed that the proposed model has significantly improved the performance of the RF, LSVM, LDA, GNB and kNN models, compared to their performance when used to segment the images alone. Furthermore, the results showed that the proposed model outperformed other models from related studies, thereby, attesting its superior segmentation capability.

Keywords: Segmentation; Machine Learning; Supervised Approach; Deep learning

Citation: Kwenda, C.; Gwetu, M.; Fonou-Dombeu, J.V. Hybridizing Deep Neural Networks and Machine learning Models for Aerial Satellite Forest Image Segmentation. *Journal Not Specified* **2023**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

1. Introduction

Forest constitute a greater portion of the natural ecosystem and is one of the richest forms of resource that contributes to the Gross National Product (GNP) of many nationalities. They play a pivotal role in areas such as climate regulation, environmental improvement, the global water cycle, and soil conservation [1][2]. Apart from a wider range of ecological services, forests contribute to socio-economic through the provision of forest product such as timber and also offers nature-based recreation [3]. Therefore it is of paramount importance to continuously monitor forests in order to understand the changes that occur with respect to time. Surveys were used to conduct forest monitoring, but such a technique is costly and cannot be completed in a short period of time [4]. With the development of advanced modern sensors, remote sensing has made it possible to monitor land cover on a large scale. However, the automatic segmentation of aerial satellite images to visualize areas that are populated with forests remains a challenging task [5].

In fact, the process of segmenting images of remote sensing of the Earth's surface has not been brought to automation with the same accuracy as with manual marking [6], despite the rapid development of computer vision algorithms for detecting objects in an image. Although humans can outperform computers in solving segmentation problems, doing so manually would take too long. Satellite image segmentation using computer vision algorithms is a very pertinent task in this scenario because it is hard to obtain segmentation results in real time. The segmentation process is generally a difficult task due to two main challenges: one is the intrinsic ambiguity in image perception and the other is that an image with many visual patterns becomes too complex to model [7].

The recent development of deep neural networks such as Convolutional Neural Networks (CNN) has improved images segmentation results [8] [9] [9]. However, deep-learning techniques produce excellent results when trained on large data sets, which is contrary to traditional machine-learning techniques which produce good results on a limited dataset. On the other hand, traditional machine learning classifiers for segmenting images such as Linear Support Vector Machines (LinearSVM), k-nearest neighbor (kNN), Linear Discriminant Analysis (LDA), and Gaussian Naive Bayes (GNB), lack the ability to extract features such as the spatial relationship between pixels and texture, resulting in subpar segmentation results when used alone. This study aims to address this shortcoming of traditional machine learning algorithms by proposing a hybrid model that combines the strengths of deep neural networks and traditional machine learning algorithms for improved segmentation of aerial satellite images. Features of aerial satellite images are firstly extracted with VGG16 and ResNet50 deep neural network models. The resulting features are subsequently used by RF, LinearSVM, kNN, LDA, and GNB classifiers to perform the final segmentation. VGG16 and ResNet50 were chosen in this study because they have innately dissimilar network architectures that abstract unrelated information for the purpose of object detection. The performance of the proposed hybrid model was evaluated using various metrics including Accuracy,Jaccard score index and Root Mean Square Error (RMSE). The aerial satellite forest images were obtained from a deep globe challenge dataset. The performance of the proposed model was evaluated using various metrics such as Accuracy, Jaccard score index and Root Mean Square Error (RMSE). The experimental results revealed that the RF model achieved the best segmentation results with accuracy, Jaccard score and RMSE of 94%, 0.913 and 0.245, respectively; followed by LSVM with accuracy, Jaccard score and RMSE of 89%, 0.876, 0.332, respectively. The LDA took the third position with accuracy, Jaccard score and RMSE of 88%, 0.834 and 0.351, respectively, followed by GNB with accuracy, Jaccard score and RMSE of 88%, 0.837 and 0.353, respectively. The kNN occupied the last position with accuracy, Jaccard score and RMSE of 83%, 0.790 and 0.408, respectively. The experimental results also revealed that the proposed model has significantly improved the performance of the RF, LSVM, LDA, GNB and kNN models, compared to their performance when used to segment the images alone. Furthermore, the results showed that the proposed model outperformed other models from related studies, thereby, attesting its superior segmentation capability.

The rest of the paper is structured as follows. Section 2 reviews related studies. An overview of feature extraction algorithms is provided in Section 3. Machine learning classifiers are discussed in section 4. Section 5 looks at the segmentation process by RF. The structure of the proposed model is presented in Section 6. Section 7 discusses the results obtained and Section 8 concludes the paper.

2. Related studies

Segmenting forest, land cover and satellite images have been of interest to many researchers [10], [11], [12], [13] in recent years. A study [10] proposed a U-net model to perform segmentation tasks on forest and water bodies satellite images. The purpose of the model was to determine the area covered by forest and water. The approach performed well as it attained a validation accuracy of 82.92% and 82.5% to perform segmentation on areas covered by water and forest respectively. This study had the challenge of having

mislabelled masks in its data set. The presence of mislabelled masks hugely contributes to the decrease in model performance. In this proposed study, the data set used did not contain any mislabelled mask. The cleaning process of removing the mislabelled mask is a very essential prepossessing task, as this has got a bearing on the overall model performance.

Another study [11] came up with a model based on U-net adopted under transfer learning to perform agricultural field segmentation on satellite imagery. The model integrated the strength of transfer learning, residual network, and U-net architecture (TL-ResUnet). The approach was tested on satellite images obtained from the DeepGlobe data set. The model outperformed other methods such as DFCNet, DeepLabv3, and DeepLabv3+ in terms of Intersection over Union (IoU). The TL-ResUnet obtained an IoU of 81%, DeepLabv3 (74.5%), and DeepLabv3+ (75.6%). However, in terms of robustness, the model failed in some circumstances to segment small forested areas and narrow water bodies because of the presence of noise in satellite images. To overcome this challenge the proposed study used a non-local means algorithm to denoise the input images.

Authors in [12] used different machine learning algorithms such as Fully convolutional neural network (FCNN), linear support vector machine, naive bayes, and logistic regression to perform semantic segmentation on satellite images to determine the allotment of forested areas in order to determine the rate at which deforestation occurs over a period of time. The FCNN achieved the highest Jaccard index score of 91.8% followed by the regression logistics with 90%. However, because of a huge imbalance in the data set the model did not perform well in detecting non-forest areas. Another challenge for FCNN was that it required a lot of training time and it consumed a significant amount of storage space.

Segmentation of forested regions in aerial images was accomplished in a study [6] using an Unet network model with 2 encoders. The model was applied to a dataset consisting of 17 images, each of which had a 16-bit channel. With a dice coefficient of 0.765, the model demonstrated its ability to segment forests in satellite images. Due to experts' inability to completely segment ground truth images, the model's detection performance was subpar, with an F-measure of 0.349. This proposed research uses a fully segmented ground truth image to overcome this challenge.

Another Model based on Unet was developed to perform the segmentation of deforestation areas using satellite images taken from Ukrainian forests [14]. Satellite images at a resolution of 512 by 512 pixels contains sections of forest, deforestration, and other areas. The dataset had an imbalance issue, however, the hybrid loss function was employed to overcome the challenge. To evaluate the effectiveness of the model as well as its consistency during the process of validation and training, k-cross validation and random runs were used. The model had an intersection over union (IOU) mean of 0.03 and an intersection over union standard deviation (IOU std) of 0.03 after running it for 100 epochs. According to these findings, the unpredictability of the initialization process and the variety of the photos did not have a major impact on the performance of the model. However, wider data variability decreased model's performance.

A study in [15] developed a supervised artificial neural network for plant image segmentation using a raw image dataset of 8-bit RGB intensity values. The neural network structure is composed of 1024 neurons in the first hidden layer, then 512 neurons in the second layer. The ReLU activation function is employed between the input layer and the first hidden layer. The sigmoid function is used between the hidden layer and the output layer. The output layer has one neuron used to compute an instance belonging to a class. The model performed well as it produced a very low error rate of 0.007. However, the approach took a significantly huge amount of time to segment images of high resolution, therefore, the study recommended using distributed computing or a graphical processing unit (GPU) to speed up the segmentation time. It is, for this reason, the proposed model in this study adopted the cloud-based GPU platform for implementation.

Another study [16] employed U-Net Neural Network with ResNet34 to conduct wildfire segmentation on satellite pictures. The adaptive moment estimation approach was utilized to achieve optimal results during the training of the model. The Resurs dataset

which is made up of 10-bit images with three channels that have a spatial resolution of between 1 and 10 meters per pixel, and the Planet dataset, which is made up of 10-bit satellite images with three channels that have a spatial resolution of 3 meters per pixel were utilized. The performance of the model was satisfactory, as it obtained a Jaccard score index of 0.87 for the Resurs dataset and 0.757 for the planet dataset. However, the application of random chromatic distortion to boost the model's robustness in the face of noisy images resulted in a minor decline in the quality of the deep learning method.

Drones, with their excellent spatial resolution and adaptability in picture capture, have just ushered in a new era in the mapping of wetlands. A study in [17] used machine learning algorithms and deep learning algorithms using drone imagery to make a map of the important plant groups in Clara Bog, an Irish wetland, before spring. The highest accuracy in semantic segmentation (about 90%) was achieved by combining the ResNet50 and SegNet architectures, and the Random forest (RF) was found to be the best pixel-based machine learning classifier. When used with the graph cut method for image segmentation, it gave good accuracy of 85%. However, the deep learning architecture's main challenge is computational overhead. To address this issue, the proposed model in this study has adopted an ensemble of VGG16 and ResNet50 models solely for feature extraction and the segmentation process is then performed by the Random Forest algorithm.

[13] implemented a random forest algorithm on SPOT satellite imagery to identify the best segmentation scales to predict land cover classes. The algorithm achieved an overall accuracy of 85.2%. The study used the Normalised Difference Vegetation Index (NDVI) and Normalised Difference Water Index (NDWI) to extract features. However, NDVI is affected by saturation, atmosphere effects, and sensor quality [18]. It is, for this reason, the proposed study adopts the hybrid approach of deep learning networks which excels at extracting features regardless of atmospheric effects. A study in [19] employed convolutional neural networks (CNN) to segment an aerial satellite image into different regions, but the study encountered challenges of having isolated satellite and segmentation images made at different times leading to inaccuracy. Another CNN segmentation model in [20] was employed to detect forest fires in aerial satellite images. Ref [21] developed a model that used NDVI to separate forest and dense grass in satellite vegetation images. Using only spectral characteristics to distinguish grass areas from forests may not yield greater accuracy, however, deploying machine-learning algorithms would yield greater accuracy for complex features.

3. Overview of feature extraction algorithms

In this study, ResNet50 and VGG16 deep learning models were used to extract features based on the studies done by [22][23]. The salient features of the models are described as follows:

3.1. VGG16 Network Model

The VGG network model was first proposed by the Visual Geometric Group at Oxford University, and that is where its name was derived. The network became much more popular in 2014 when it won first and second place in the classification and localization task when it participated in the ImageNet Large Scale Recognition Challenge (ILSVRC) [24]. The network is composed of 13 convolutional layers and 3 fully connected layers, hence the name VGG16. The large convolution filter in the VGG16 network has been replaced by various 3×3 convolutional filters stacked on top of each other. The multiple 3×3 convolutional filters make the network deeper and at the same time reduce the number of total parameters [25]. In the VGG16 network, each max-pooling layer has a kernel of size 2 and a step of 2.

3.2. ResNet50 Network Model

ResNet50 is a 50-layer-deep CNN and is the first network to adopt residual learning in 2015 [26]. The network won the first prize in 2015 when participated in computer vision

benchmarking challenges in the ILSVRC and Microsoft Common Objects in Context2015. A deep layered network suffers from increased error rates due to the vanishing gradient problem [27]. However, the ResNet50 models solve this challenge by incorporating a technique called skip connections or shortcuts as shown in Figure 1. The shortcuts jump several layers and connect directly to the input, hence the vanishing gradient is avoided. The mapping function of a shortcut connection sums up the input instance and the output instance such that the original mapping function

$$H(x) = F(x) - x \tag{1}$$

is redefined as

$$H(x) = F(x) + x \tag{2}$$

The refinement of the mapping function makes learning simple whilst the desired functionality is achieved. The mapping function presented in equation 2 is implemented through feed-forward neural networks as presented in Figure 1

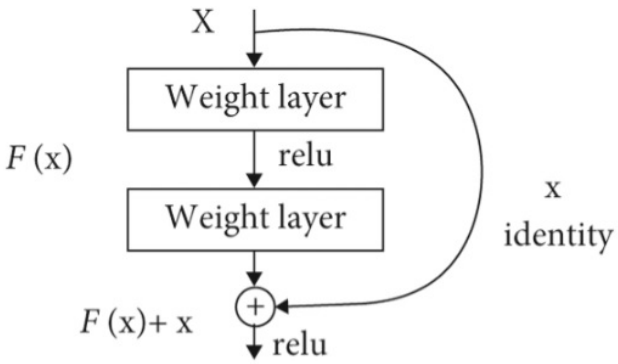


Figure 1. ResNet50 shortcut

4. Overview of Machine Learning Classifiers

4.1. Random Forest Algorithm

The random forest algorithm is an ensemble classifier based on decision trees, where each tree grows through randomization. The RF algorithm is capable of processing large amounts of data at high speed using decision trees. During the training phase, the RF algorithm randomly chose a subset of data from the training data. At a particular node, say n, the training data is recursively split into left and right subsets using the split function and the threshold. The split function randomly selects the threshold in the range $h \in (minX(v_i), maxX(v_i)$ where h is the threshold and $X(v_i)$ is the split function of vector v. The split function that creates the left and right subset trees is expressed as:

$$m_l = (i \in m_n)|x(v_i) < h) \tag{3}$$

$$m_r = m_n \backslash m_i \tag{4}$$

where m_l is left data, m_r is right data and m_n is data at corresponding node n. At the split node, several candidates are randomly produced through the split function and the threshold. Only candidates that maximize the information gain at a given node are selected. The information gain is computed by entropy estimation as expressed in equation 11.

$$\Delta E = -\frac{|m_i|}{|m_n|}E(m_n) - \frac{|m_r|}{|m_n|}E(m_n) \tag{5}$$

where ΔE is the information gain. Whenever the training process reaches a leaf node or no more ΔE is possible, the iterative process stops. The final class is generated by the ensemble of all distributed trees X = $(x_1, x_2, ..., x_n)$ as presented in equation 16:

$$P(c_i|X) = \frac{1}{N}\sum_{n=1}^{N} P(c_i|x_N) \tag{6}$$

where $P(c_i|X)$ is the probability of class c_i given distributed trees X.

4.2. Linear Support Vector Machines

Linear Support Vector Machines (LinearSVM) is a machine learning classification technique that was proposed by Vapnick and his group at AT&T BELL laboratories [28][29]. LinearSVM works on obtaining the best generalization performance by ensuring a relationship balance between accuracy obtained from the training data and the machine capacity. It works by trying to separate classes with a hyperplane surface so as to maximize the margin among them. LinearSVM has also been successfully applied to perform handwritten digit recognition, face detection on images, and object detection [30]. Based on Vapnick, LinearSVM can either be described either from the linearly separable case or Non-linearly separable case.

4.2.1. Linearly seperable case

For this case, data is considered to be linearly separable, and the plane is defined by an equation: $v.x + c = 0$, where x is a specific point on a hyperplane, and v is an m-dimensional vector that is perpendicular to the hyperplane, and c is the distance of the point is closest to the hyperplane origin. This will arise two inequality equations:

$$v.x_i + c \geq 1, \quad for \;\; y_1 = +1, \;\; and \tag{7}$$

$$v.x_i + c \leq 1, \quad for \;\; y_1 = -1 \tag{8}$$

Equation 7 and 8 can be combined into

$$y_i(v.x_i + c) - 1 \geq 0, \forall i \tag{9}$$

Now LinerSVM will try to find a hyperplane $v.x + c = 0$ with minimum $||v||^2$. This is also equivalent to determining the hyperplane with the largest margin, which is determined by calculating the distance between the closest vectors for two classes. Hence the problem is redefined as follows:

$$Minimize \underbrace{_{v,c}} \frac{1}{2}||v||^2 \;\; subject \;\; to \;\; y_i(v.x_i + c) - 1 > 0 \tag{10}$$

4.2.2. Non-Linear Case

For this case, data appears in the optimization problem in the form of dot products. It maps features vectors to a higher dimensional Euclidean space by a mapping:

$$\Phi : R^d \alpha H \tag{11}$$

Then the optimization problem in space L is obtained by replacing $x_i.x_j$ by $\Phi(x_i).\Phi(x_j)$. If there is kernel function Q defined by

$$Q(x_i, x_j) = \Phi(x_i).\Phi(x_j) \tag{12}$$

then there is only a need to compute $Q(x_i, x_j)$ in the training maps. The decision function then becomes

$$f(x) = sign(\sum_{i=1}^{I} y_i \lambda_i Q(x, x_j) + c) \tag{13}$$

4.3. k Nearest Neighbor

The K nearest neighbor algorithm is another machine learning technique that is employed for regression and classification-related tasks. It works by assigning unmarked data points to the class that is nearest to the labeled data point [31]. The algorithm's efficacy is through its ability to leverage similarity metrics, which consider the distance between points to determine the most analogous data point. K-nearest neighbor applies information obtained from the observed data to make its predictions rather than relying on predefined associations between the predictor and predicted variable. Fore regression related tasks, the K-NN approximates the response of a test point (x_p) by considering the weighted average of all the responses from the closest point $(x(1), x(2), ..., x(k)$ in the vicinity of (x_p). In order to determine the right weight to assign to each neighbor; kNN adopts a kernel function that calculates the weight of the neighbor based on its proximity to the test point. For a given training dataset $X = \{x_1, x_1, ..., x_s\}$ consisting of s training points, each with T features, weighted Euclidean distance can be employed to determine the distance between each training point xi and the test point (x_p). Euclidean Distance (ED) is computed as presented in the equation 14

$$ED(x_p, x_i) = \sqrt{\sum_{t}^{T} w_t (x_{p,t} - x_{i,t})^2} \tag{14}$$

where T represents the number of features, $x_{p,t}$ denotes the t^{th} feature value of the existing point x_p, $x_{i,t}$ denotes the t^{th} feature value of training point x_i. The t^{th} feature weight is represented by w_t. The kernel regression that is used to estimate the response of x_p is defined in equation 15

$$f(x_p) = \frac{\sum_{i=1}^{k} \phi(x_p, x_i) f(x_i)}{\sum_{i=1}^{k} \phi(x_p, x_i)} \tag{15}$$

where k is the number of k-nearest neighbors, $\phi(x_p, x_i) f(x_i)$ is the kernel function at the i^{th} training point and $f(x_i)$ is the known response of x_i.

4.4. Linear Discriminant Analysis

The Linear Discriminant Analysis is employed to decide to differentiate between input patterns [32]. For a given two classes the decision boundary is defined as:

$$d(Q) = Q_1 - mQ_2 - r \tag{16}$$

where Q_1 and Q_2 represents input patterns. The idea behind LDA is to construct a decision surface such that $d(Q) > 0$ would categorize patterns for one class and $d(Q) < 0$ would categorize patterns for another class. Considering that $x = \{x_1, x_2, ..., x_M\}$ is an M dimensional pattern vector. Suppose the number of classes is n and the problem is to classify a given instance x to any one of the classes. The problem is solved by defining n decision functions given by $d_1x, d_2x, ..., d_nx$. The instance x would be categorized into class p and not q if

$$d_p(x) > d_q(x) \;\; wherep \neq q \;\; if \;\; p, q = 1, 2, 3, ..., n \tag{17}$$

The decision boundary between the two classes p and q will be redefined as

$$d_p(x) - d_q(x) = 0 \tag{18}$$

Therefore the instance x would be classified into class p if

$$(d_p(x) - d_q(x)) > 0 \tag{19}$$

and to class q if

$$(d_p(x) - d_q(x)) < 0 \tag{20}$$

4.5. Gaussian Naive Bayes

Gaussian Naive Bayes simplifies learning by assuming that features are independent of given classes [33]. This assumption is described by many researchers as poor in general, but however, it works effectively based on this assumption. The Bayesian Classifier assigns a given instance x to the most likely class as expressed in the equation 21

$$P(C) = \prod_{i}^{n} = Ip(X_i|c) \tag{21}$$

where C denotes the classifier, and $X = (X_1, ..., X_n)$ represents a feature vector [34]. The Gaussian Naive Bayes is a simplified version of Bayesian probability based on the independence assumption. This implies that one attribute of the probability of one has no impact on the probabilities of the other attributes.

5. Evaluation metrics used in the study

Metrics such as the Jaccard index, Root Mean Square Error (RMSE), confusion matrix, ROC_AUC curves, Precision, Recall, F1-Score, and Accuracy are used in this study to evaluate the performance of the proposed segmentation model. The confusion matrix facilitates the visualization of the model's performance. The visualization platform makes it simple to identify confusion between regions, e.g., it is simple to determine which regions have more misclassified pixels than others. The ROC_AUC curve, also known as the sensitivity measure, is a graph of the true-positive rate versus the false-positive rate. Better classification performance is indicated by a model with a trajectory that is located far from the median. In a plot, the ROC_AUC curve represents the efficacy of a model across all thresholds. The bigger the area, the better the model. One of the benefits of the ROC_AUC curve is that it facilitates the comparison of results from various models without the need to reconcile sensitivity and specificity concerns. The ROC_AUC formula for binary classification is expressed in equation 22 [35].

$$ROC\_AUC = \frac{x_p - m_p(m_p + 1)/2}{m_p m_m} \tag{22}$$

where x_p denotes the sum of all positive ranked samples. m_p and m_m represent the number of negative and positive samples, respectively.

The Jaccard index, also referred to as the Intersection-Over-Union(IoU) is the widely used metric for evaluating the predictions of segmentation models. IoU is defined by the area of overlap between the predicted segmented image and the reference image(ground truth) divided by the union area of the segmented image and the reference image. The IoU is defined in Equation 23.

$$IoU = \frac{TP}{TP + FP + FN} \tag{23}$$

where TP denotes true positive, TN represents true negative, FN represents false nagative and FP denotes false positive. Accuracy determines the efficiency of the model by considering the total correct predictions made by the segmentation method. Accuracy is expressed in Equation 24.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{24}$$

The root-mean-square error RMSE is the square root of the mean square of all errors. Because it is scale-dependent, RMSE is a good measure of accuracy for comparing forecasting errors of different models or model configurations for a specific variable but not between variables. It is calculated in Equation 25.

$$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^{n} (O_i - P_i)^2} \tag{25}$$

where O_i are the actual values and P_i are the predicted value Recall also referred to as sensitivity is a measure of instances predicted as positive against all actual positive values. This metric returns the fraction of positive patterns that are correctly classified. Recall metric is computed by equation 26

$$Recall = \frac{TP}{TP + FN} \tag{26}$$

Precision returns the proportion or fraction of positive identification (true positives) that were correct. Precision is expressed in equation 27.

$$Precision = \frac{TP}{TP + FP} \tag{27}$$

F1- Score is the harmonic average between precision and recall rates. This metric is expressed in equation 28

$$F1 - Score = 2x\frac{PrecisionxRecall}{Precision + Recall} \tag{28}$$

6. Materials and methods

6.1. Data set

The aerial image used for the study was obtained from Land Cover Classification Truck in the DeepGlobe Challenge data set [36]. The associated reference image in the data set is binary in nature, it only shows forest regions areas and non-forest regions areas. Figure 2 shows an original image and its corresponding labeled mask from the dataset. The experiment was conducted on the Google Colab environment which provides free GPU and TPU cloud resources. In particular, the experiment used GPU with the acceleration of NVIDIA Tesla due to the high computational requirements of the experiment.

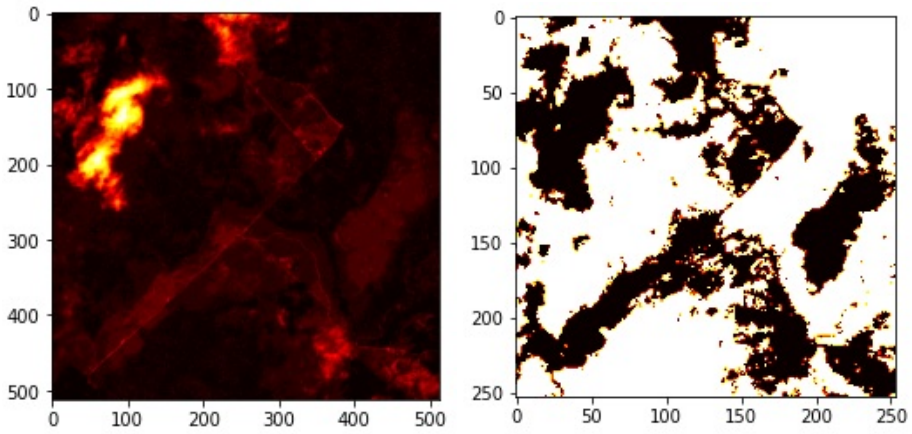


Figure 2. The right panel shows the extracted RGB patches and the left panel shows its corresponding masks.

6.2. The proposed Model

The study proposes an aerial forest image segmentation model that uses a hybrid approach of ResNet50 and VGG16 deep learning models to generate a set of features for the machine learning algorithms to perform the segmentation process. This study chose these two pre-existing models due to their innately dissimilar architecture that abstracts unrelated information from images used for object detection purposes [37]. The hybrid approach helps in expanding the feature vector scope. A single feature selection method only chooses the best subset of features from the training dataset. As a result, the end feature vector may not be a true reflection of the training dataset and may not be a good

starting point for the next step, which is to segment the image. When different methods' results are put together, the result may be more accurate. Features produced by the hybrid approach of deep learning models were applied to various traditional machine learning techniques such as K Nearest neighbor, Random Forest, Linear Discriminant Analysis, Gaussian Naive Bayes, and Linear Support Vector Machines to evaluate their segmenting power on aerial satellite forest test image. The general framework of the model is shown in Figure 3.

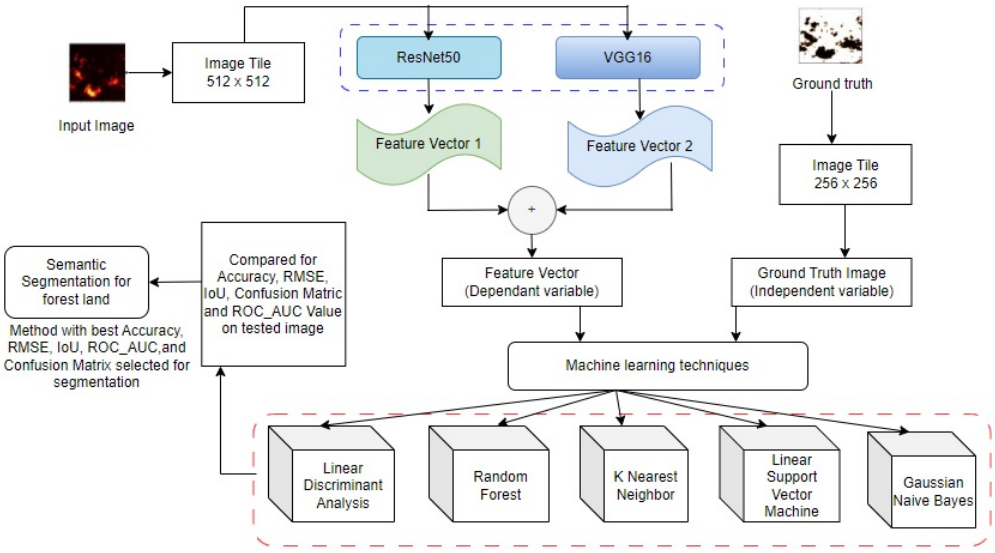


Figure 3. Segmentation framework model with all traditional machine learning classifiers

6.2.1. Feature Generation

The VGG16 and ResNet50 deep-learning models were used to generate a set of features which were in turn used to segment an aerial forest image. VGG16 managed to produce 128 features (Figure 4) and ResNet50 managed to produce 64 features (Figure 5). The output features from both models were concatenated together to produce the final feature vector with 192 features (Figure 6).

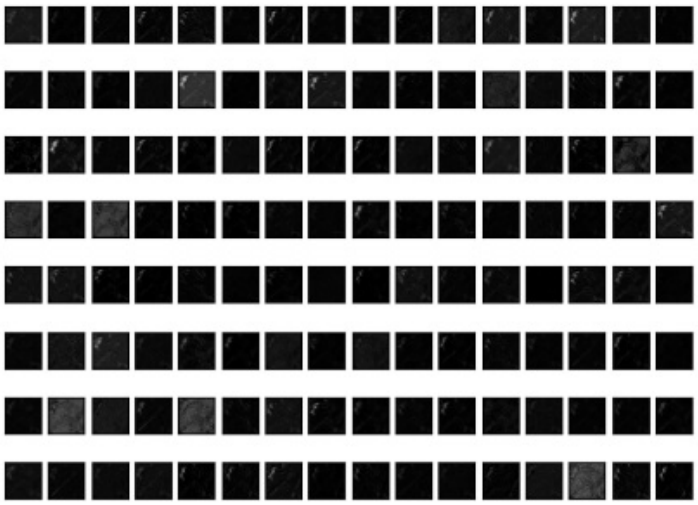


Figure 4. set of features produced by VGG16 model

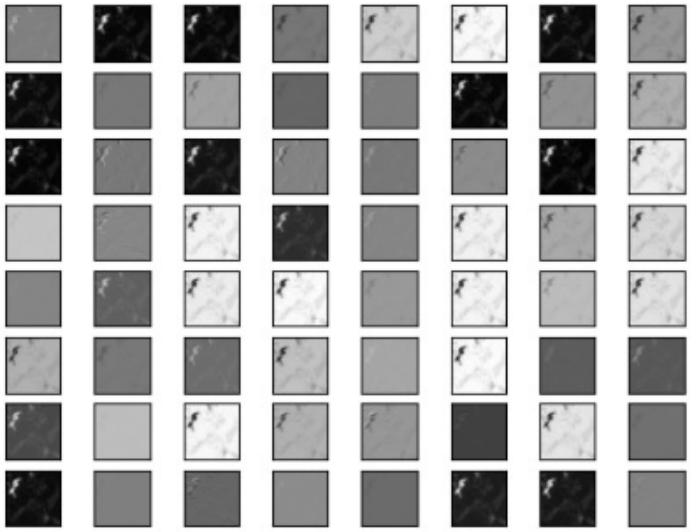


Figure 5. set of features produced by ResNet50 model

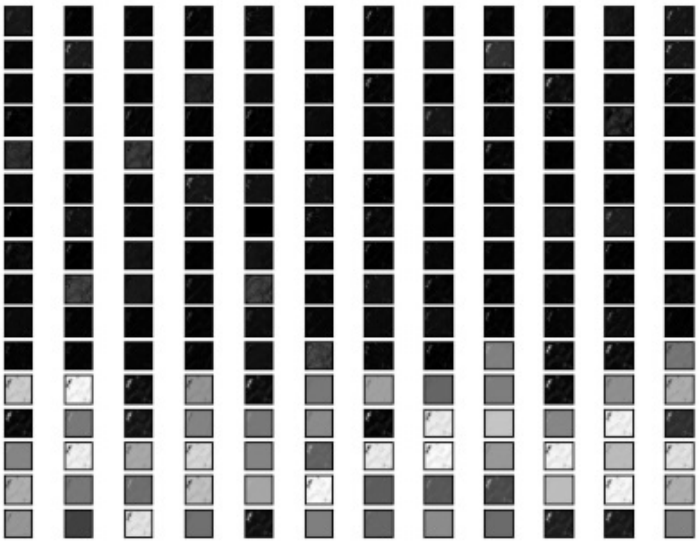


Figure 6. composite set of features produced by VGG16 and ResNet50

6.2.2. Segmentation process by machine learning classifiers

Combined features obtained from the hybridized approach of VGG16 and ResNet50 models are set as a dependent variable in the data frame. The pixel values obtained from the ground truth image are also set as an independent variable in the data frame. As presented in Algorithm 1, X is a vector that contains all the features extracted by VGG16 and ResNet50. These sets of features are set as independent variables. Variable M contains ground truth image pixel values which in turn are set as dependent variables. These two sets are split into a train set and a test set and the machine learning classifier is adopted to predict the segmented image.

7. Results and Discussion

As explained in the proposed model, the set of features generated by a hybrid approach of VGG16 and ResNet50 models is provided as an input to various machine learning classifiers. The goal is to determine the type of classifier that produces a satisfactory result. In the following subsections, the performance of these models was evaluated in terms of F1-score, precision, and recall for detecting forest areas and non-forest areas.

Algorithm 1 An algorithm for machine learning classifier segmentation

Input : P(y): Ground_truth_pixel_ values
Input : P(x): Independent_variables_pixel_values
for `P(x) = 0` **do** ▷ All features generated must match how features are generated for training
    $feature1 \leftarrow VGG16$
    $feature2 \leftarrow ResNet50$
    $feature3 \leftarrow ground\_truth\_pixel\_values$
end for
$X \leftarrow \sum_1^2 feature(x)$ ▷ Features are added to the data frame
$M \leftarrow feature3$ ▷ M denotes an independent variable
X ⫫ M
Input : Data: Train_set ▷ New Train Set from the extracted feature now to be loaded to classifier
Data = train + test ▷ Test data for accuracy testing
model = MachinelearningClassifier()
model.fit(X_train, M_train)
prediction_test = model.predict(X_test)
loaded_model = pickle.load(open(filename, 'rb')) ▷ Applying trained model to segment other images
Return segmented_image

7.1. Evaluation of machine learning models in detecting forest region areas

Table 1 presents the performance of each machine learning classifier in detecting forest regions in terms of F1 score, precision, and recall. The Random Forest algorithm attained the highest precision of 0.94, followed by LinearSVM, LDA, GNB, and kNN with precision scores of 0.92, 0.88, 0.86, and 0.82. On the other hand, GNB, obtained the highest recall of 0.98, outperforming the other machine-learning classifiers. Again, the RF algorithm recorded the highest f1-score compared to the other algorithms.

Table 1. Metrics of classifiers in terms of precision, recall, and F1-Score with a hybrid approach of deep learning for detecting forest areas.

Metric	kNN	RF	LDA	LinearSVM	GNB
Precision	0.82	0.94	0.88	0.92	0.86
Recall	0.95	0.96	0.95	0.95	0.98
F1-Score score	0.88	0.94	0.88	0.91	0.88

7.2. Evaluation of machine learning models in detecting non-forest region areas

Table 2 presents the performance of each machine learning classifier in detecting non-forest regions using the same metrics of F1 score, precision, and recall. The Linear Discriminant Analysis technique attained the highest precision of 0.94, followed by RF, LinearSVM, LDA, and kNN with precision scores of 0.93, 0.90, 0.88, and 0.87. On the other hand, RF, obtained the highest recall of 0.89, outperforming the other machine-learning classifiers. Again, the RF algorithm recorded the highest f1-score compared to the other algorithms.

Table 2. Metrics of classifiers in terms of precision, recall, and F1-Score with a hybrid approach of deep learning for detecting non-forest areas.

Metric	kNN	RF	LDA	LinearSVM	GNB
Precision	0.87	0.93	0.88	0.90	0.94
Recall	0.61	0.89	0.75	0.84	0.69
F1-Score score	0.72	0.91	0.81	0.87	0.79

7.3. Evaluation of machine learning models in segmenting aerial satellite forest image.

The power of segmenting an aerial forest image for each machine learning was evaluated in terms of RMSE, accuracy, and Jaccard score. As presented in Table 3, the model based on RF outperformed other machine learning segmentation techniques such as Gaussian Naive Bayes (GNB), k Nearest Neighbor (kNN), Linear discriminant analysis (LDA), and Linear Support Vector Machine (LinearSVM) in terms of accuracy, Jaccard score index, and RMSE. In terms of errors, the RF-based model recorded the lowest RMSE of 0.245, and this implies that its predictions are much closer to the actual values than those of other models. Again, the RF-based model also achieved the highest IOU of 0.913, indicating a minimum overlap between the target mask and the predicted output, and also the same technique had the highest accuracy of 0.94 implying that most pixels were classified into their true regions. Figure 7 shows segmented images of the test image with respect to all the classifiers used in a hybrid deep learning approach.

Table 3. Metrics of classifiers in terms of accuracy, RMSE, and Jaccard score with a hybrid approach of deep learning in segmenting aerial satellite forest image.

Metric	kNN	RF	LDA	LinearSVM	GNB
RMSE	0.408	0.245	0.351	0.332	0.353
Accuracy	0.833	0.940	0.877	0.890	0.876
Jaccard score	0.790	0.913	0.834	0.876	0.837

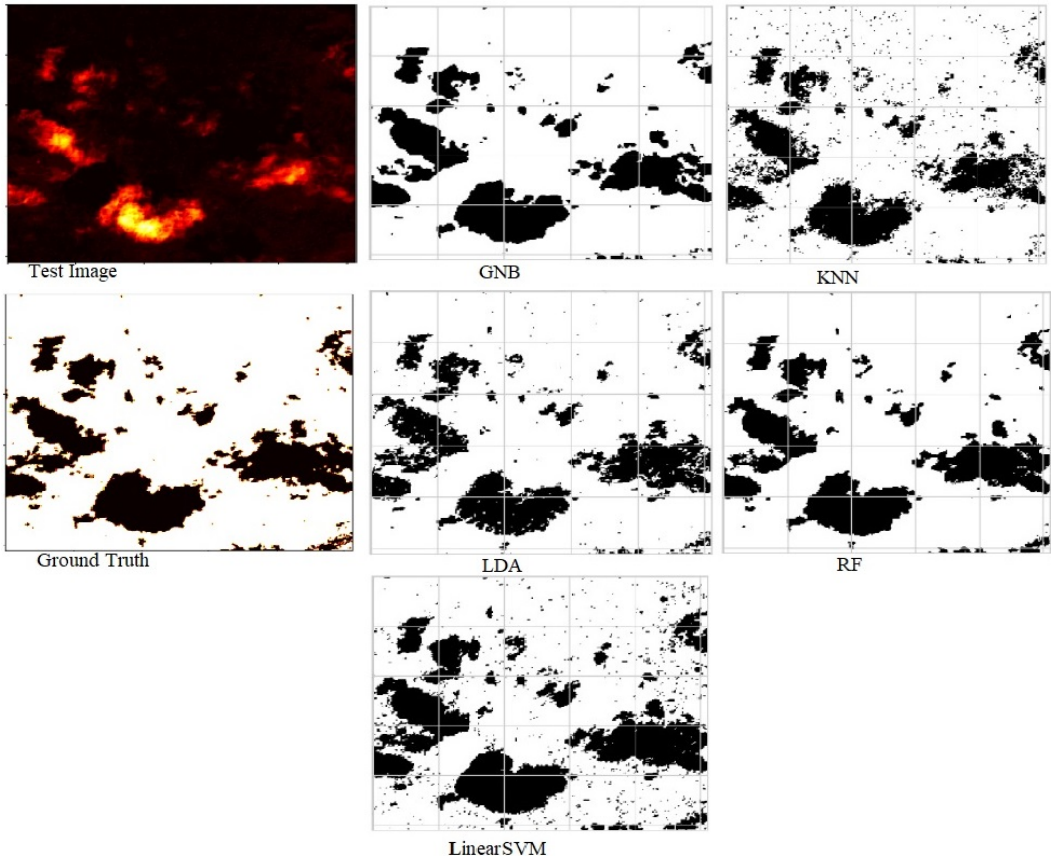


Figure 7. Predicted segmentation results by RF, LDA, GNB, kNN, and LinearSVM with the ensembled approach

Figure 8 presents the confusion matrix of all the classifiers in response to the features obtained from the hybrid approach of deep learning models. The Gaussian Naive Bayes confusion managed to classify 8375 of the 8589 pixels in class 1, while 214 were misclassified as belonging to class 0. Only 1418 pixels were misclassified as class 1 out of 4519 pixels

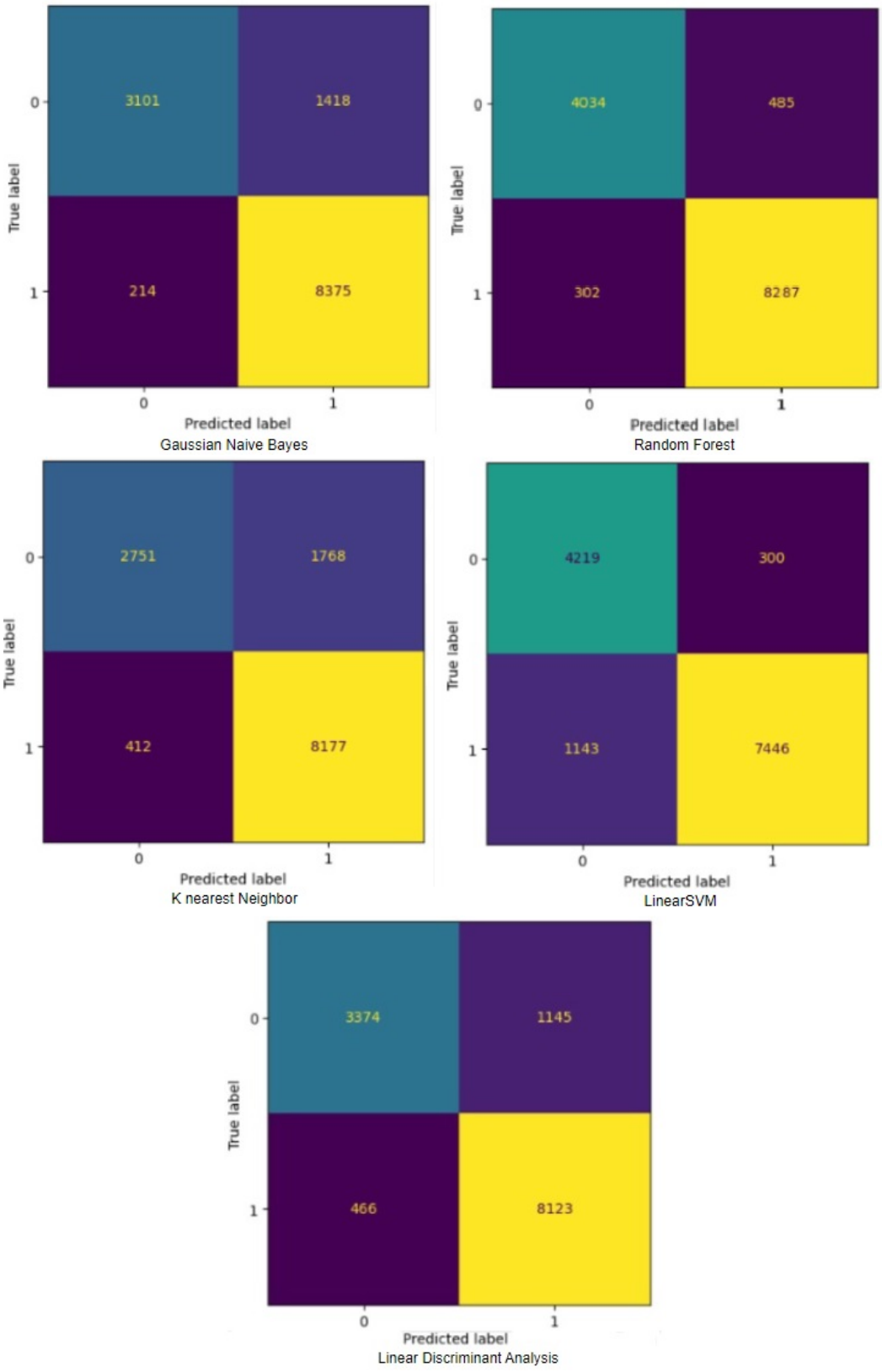


Figure 8. Confusion matrix results for RF, LinearSVM, LDA, GNB, and kNN

classified as class 0. The model misclassified 1632 pixels in total. The Random Forest-based model recorded the best performance in relation to the other 4 models. The model recorded the least pixel misclassification compared to the other models as it misclassified

only 787 pixels. Linear Support Vector Machine confusion matrix indicates that, out of 8589 pixels that belong to class 1, 8123 were correctly classified and 466 were misclassified as belonging to class 0. For 4519 pixels under class 0, 3374 were correctly classified and 1145 were misclassified as class 1. In total, the model misclassified 1611 pixels. The LDA model correctly classified 11497 pixels and wrongly classify 1611 pixels to other classes. The model based on kNN performed the least of all the other algorithms as it misclassified most pixels into other classes. The ROC_ AUC curves in Figure 9 show that all the classifiers have excellent potential to distinguish regions in an image as indicated by their values that are above 0.9. The model based on Random Forest emerged as the best model at distinguishing image objects as it attained the highest ROC_ AUC value of 0.98.

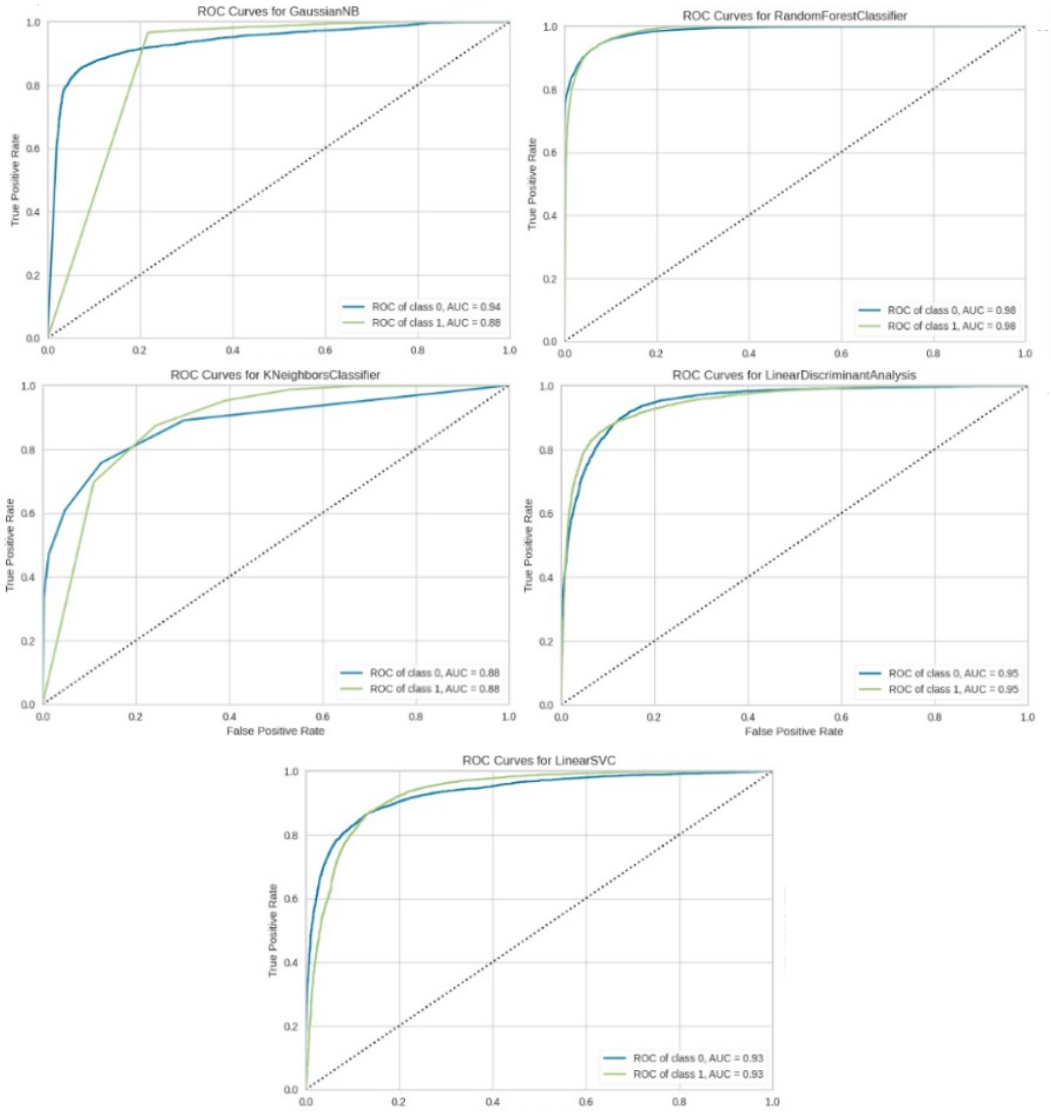


Figure 9. ROC curves for RF, GNB, LDA, LinearSVM, and kNN

7.4. Evaluation of the performance of the classifiers without the hybrid deep learning approach in detecting forest areas

Table 4 shows the performance of the machine learning classifiers without the hybrid deep learning approaches in detecting forest region areas in terms of precision, recall, and F1 score. The RF, LDA, LinearSVM, and GNB obtained the same precision score of 0.65 with kNN obtaining a slightly low score of 0.64. Again RF, LDA, LinearSVM attained an absolute recall value of 1.0 with kNN performing the least with a recall value of 0.56. Regarding F1-score, kNN performed the least by obtaining an F1 score of 0.49.

Table 4. Metrics of classifiers in terms of precision, recall, and F1-Score without the deep learning hybrid approach in detecting forest areas.

Metric	kNN	RF	LDA	LinearSVM	GNB
Precision	0.64	0.65	0.65	0.65	0.65
Recall	0.49	1.00	1.00	1.00	0.98
F1-Score score	0.56	0.78	0.78	0.78	0.78

7.5. Evaluation of the performance of the classifiers without the hybrid deep learning approach in detecting non-forest areas

Table 5 also shows the performance of the machine learning classifiers without the hybrid deep learning approach in detecting non-forest regions. The RF algorithm had the best precision score of 0.50, while LDA and LinearSVM had the lowest performance with precision values of 0. The kNN approach had the maximum recall value of 0.50, whereas LinearSVM and LDA had the lowest recall values of 0. In terms of F1 score, the kNN machine learning technique outperformed all other classifiers once more, with a value of 0.41

Table 5. Metrics of classifiers in terms of precision, recall, and F1-Score without the deep learning hybrid approach in detecting non-forest areas.

Metric	kNN	RF	LDA	LinearSVM	GNB
Precision	0.35	0.50	0.00	0.00	0.44
Recall	0.50	0.01	0.00	0.00	0.03
F1-Score score	0.41	0.01	0.00	0.00	0.05

7.6. Evaluation of machine learning models without the hybrid deep learning approach in segmenting aerial satellite forest image.

To further evaluate the performance of the five models, we computed the RMSE, Accuracy, and Jaccard score index for each classifier. Table 6 presents the RMSE, Accuracy, and Jaccard score for each classifier. It is shown in Table 6 that the accuracy and Jaccard score of RF, GNB, LDA, and LinearSVM comparatively obtained the same value of around 0.65 which indicates an average performance. The kNN model recorded the highest error of 0.71 in terms of RMSE and overall performed the worst against all other classifiers. Figure 10 shows segmented images of the test image produced by all the classifiers without the deep learning approach.

Table 6. Metrics of classifiers in terms of accuracy, RMSE, and Jaccard score without the hybrid approach of deep learning in segmenting aerial satellite forest image.

Metric	kNN	RF	LDA	LinearSVM	GNB
RMSE	0.711	0.595	0.595	0.595	0.598
Accuracy	0.495	0.646	0.646	0.645	0.643
Jaccard score	0.386	0.645	0.646	0.646	0.639

The ROC_AUC values obtained in Figure 12 indicate that LinearSVM, LDA, GNB, and RF cannot adequately distinguish between image regions because the obtained ROC_AUC values lie in between 0.5 to 0.7. A ROC_AUC value range between 0.5 to 0.7 means that the model cannot adequately distinguish between image regions range between 0.7 to 0.8 its acceptable discrimination, 0.8 to 0.9 offers good discrimination and values that are greater than 0.9 have excellent discrimination [38]. kNN model alone is recommended to be used in object detection as it obtained a value that is less than 0.5.

Table 7 shows a comparison in performance between the machine learning classifiers with the hybrid deep learning approach and the classifiers alone in terms of accuracy, RMSE, and Jaccard index score. The classifiers alone were completely outclassed by those

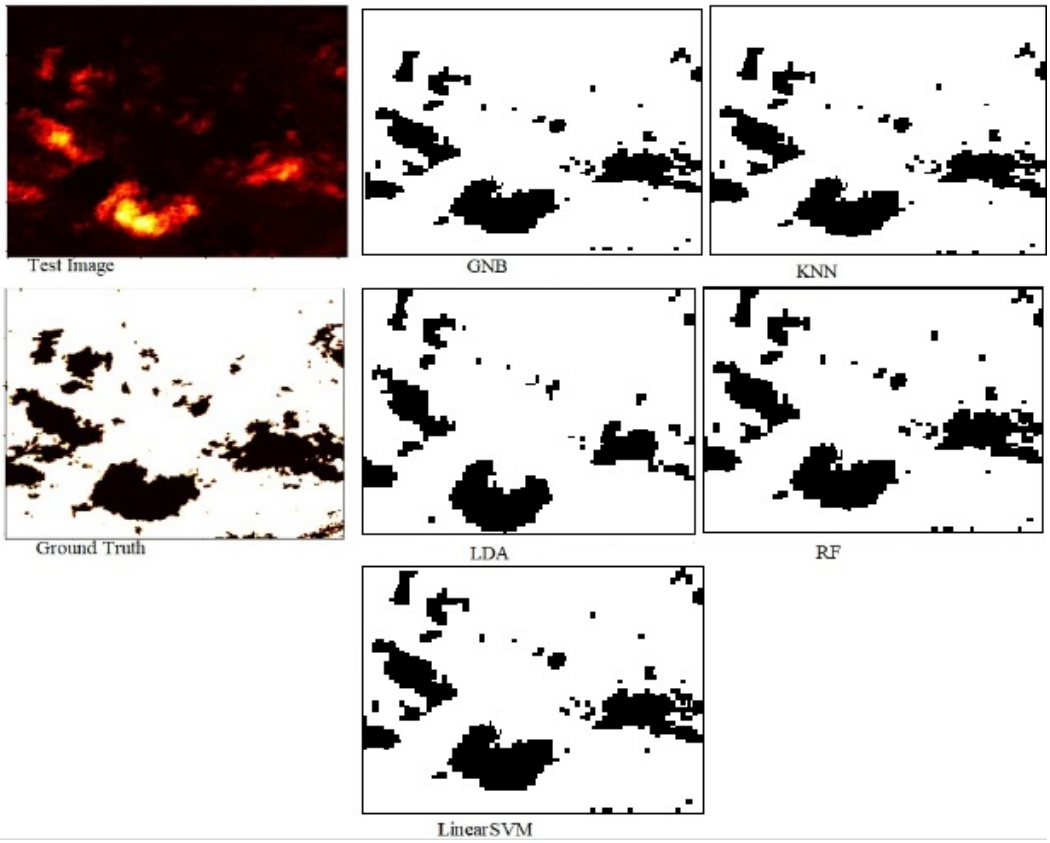


Figure 10. segmentation of test image with respect to RF, LinearSVM, LDA, GNB, and kNN without the deep learning

Table 7. Performance of the classifiers with the hybrid deep learning approach vs classifiers alone in terms of RMSE, Accuracy, and Jaccard score index

Classifiers with hybrid deep learning approach						**Classifiers Alone**				
Metric	**RF**	**LinearSVM**	**LDA**	**GNB**	**kNN**	**RF**	**LinearSVM**	**LDA**	**GNB**	**kNN**
Accuracy	0.940	0.890	0.877	0.876	0.833	0.646	0.646	0.646	0.643	0.49
Jaccard Score	0.913	0.876	0.834	0.837	0.790	0.645	0.646	0.646	0.639	0.386
RMSE	0.245	0.332	0.351	0.535	0.408	0.595	0.594	0.595	0.600	0.711

with the hybrid deep learning approach in terms of all the metrics. This is attributed to the fact that the classifiers alone do not have the capacity to extract features such as spatial relationships between pixel and texture. Therefore classifiers should be used in a pipelined fashion where they perform the process of segmentation after receiving features from other models. This is the reason why the classifiers used in conjunction with deep learning hybrid approaches produced satisfactory results. In the hybridization approach, the RF algorithm emerged as the winner in performing the segmentation task. These results also go in glove with results obtained by [39] where the Random Forest approach outperformed other algorithms such as Gentle AdaBoost (GAB), Maximum Likelihood Classification (MLC), and Support Vector Machines (SVM) in a pipelined approach fashion. Another study by [40] demonstrated that the RF algorithm performs well in object detection for multi-spectral images.

It is important to evaluate segmentation algorithms to determine algorithms suitable for a given application. Algorithm performance is dependent on the type of images used. Images are generally classified into, synthetic, remote sensing, medical and natural images. A particular algorithm might be better in remote sensing images but poor in medical images. In light of the evaluation of the algorithms, the Random Forest algorithm outperformed Linear Discriminant Analysis, Gaussian Naive Bayes, and Support vector

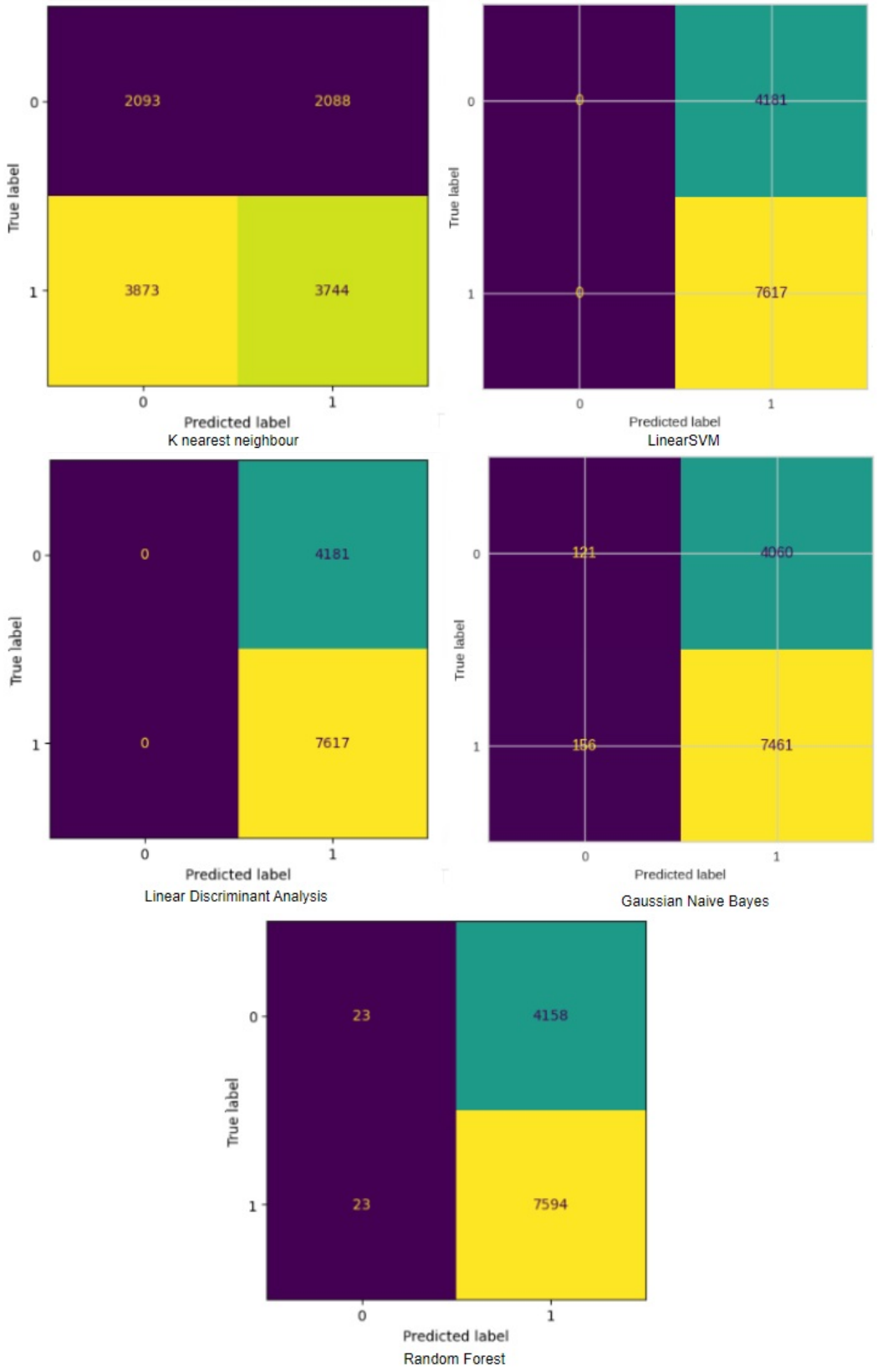


Figure 11. Confusion matrix results for RF, LinearSVM, LDA, GNB, and kNN without the deep learning

machines algorithms in terms of accuracy, Jaccard score, and ROC curves. As presented in Table 8 the proposed model in this study outperformed other models from related

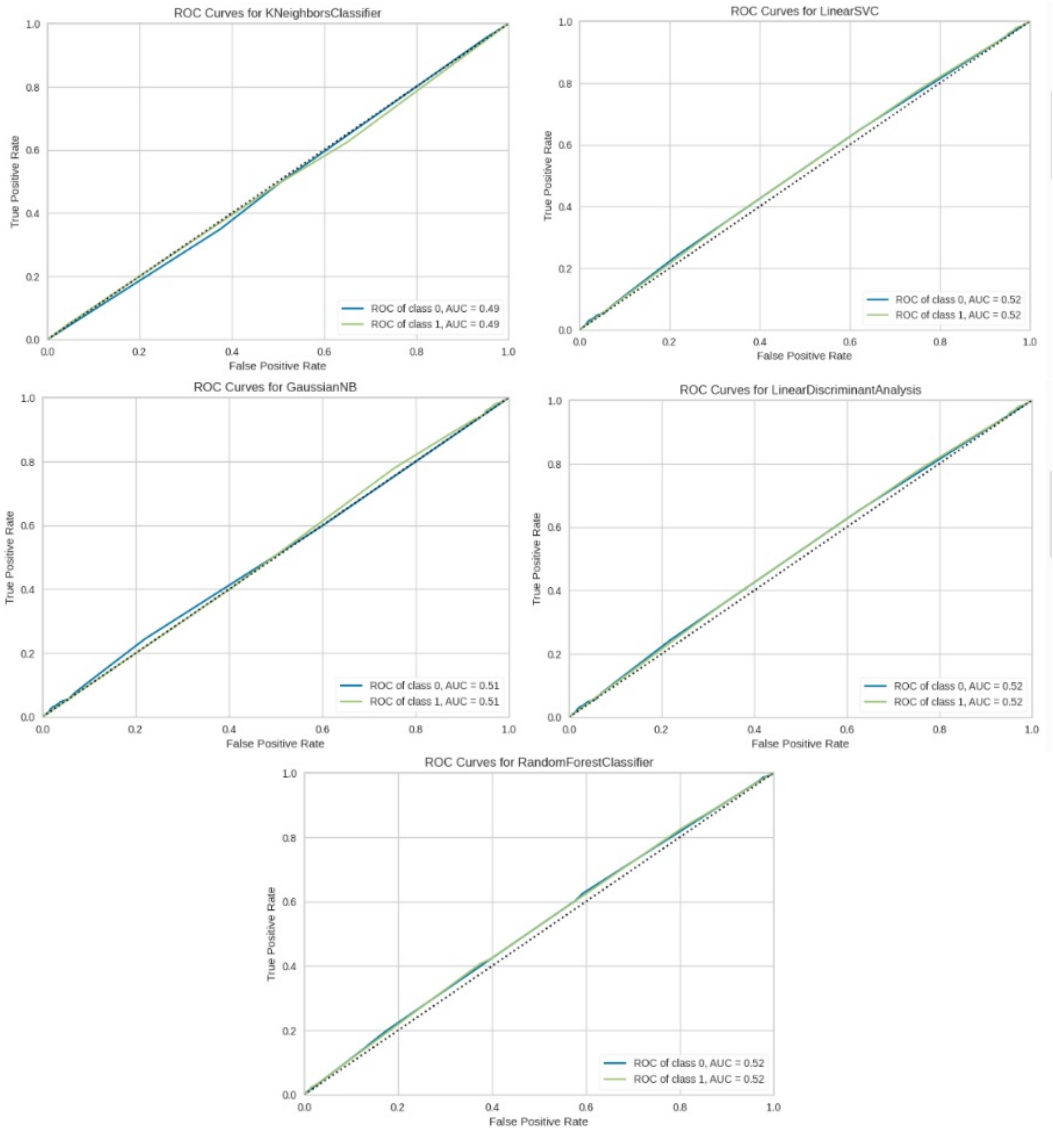


Figure 12. ROC Curve results for RF, LinearSVM, LDA, GNB, and kNN without the deep learning

Table 8. Accuracy and IOU obtained from other studies

Method	**Accuracy**	**IOU**
Unet with spatial pyramid spooling [41]	86.71	75.59
Hnet with Inception as backbone [42]	68	83
SemisFsNet [43]	-	80
Unet for forest segmentation [44]	82.55	54
Unet for forest segmentation [44]	82.95	60
improved tuna swarm optimization (ITSO) [43]	-	59
Unet semantic segmentation [45]	99	97
Random Forest	**94**	**91**

studies [41][46][44][42] [43][47]. However, the Unet semantic segmentation in [45] for Forest Change Detection in South Korea Using Airborne Imagery outperformed our model with 99% accuracy and 97% IOU. The reason could be attributed to the ability of Unet to extract more features required to perform subsequent segmentation.

8. Conclusion

This paper adopts a hybridized approach of deep learning models and traditional machine learning classifiers used to identify forest and non-forest areas from an aerial satellite image obtained from the Deep Globe challenge dataset. A deep learning hybrid approach of VGG16 and ResNet50 was used to extract a set of features that were subsequently used by machine learning classifiers to segment an aerial satellite image into the forest and non-forest areas. Metrics such as IoU, accuracy, RMSE, and ROC_AUC curves were used to assess the performance of the models. The model based on RF emerged as the winner as it achieved an accuracy of 94% and an IoU of 91%. The high efficacy of the model implies that the model can be used to detect smoke, veld fires, and perform water segmentation. The ensemble edge vector approach contributed to the high efficacy of the model. For future work, it is recommended to include more classes and to adopt high-resolution networks (HRnets) as an alternative to VGG16 and ResNet50 because of their ability to perform low-resolution to high-resolution conversion, which is also linked to their block network architectures constructed according to new standards, and therefore excels at vision tasks such as feature extraction and object detection.

Author Contributions

Introduction and related work was done by J.F. Model design was done by M.G. Implementation and discussion section was done by C.K.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable

Institutional Review Board Statement

Not applicable

Informed Consent Statement

Not applicable

Data Availability statement

The data that support the findings of this study are available on [36].The authors confirm that the data supporting the findings of this study are available within the article.

Acknowledgements

The authors thank the University of KwaZulu Natal for providing financial assistance in accessing all resources and tools required to undertake this study.

References

1. Xiao, J.L.; Zeng, F.; He, Q.L.; Yao, Y.X.; Han, X.; Shi, W.Y. Responses of forest carbon cycle to drought and elevated CO2. *Atmosphere* **2021**, *12*, 212.
2. Shaheen, H.; Khan, R.W.A.; Hussain, K.; Ullah, T.S.; Nasir, M.; Mehmood, A. Carbon stocks assessment in subtropical forest types of Kashmir Himalayas. *Pak. J. Bot* **2016**, *48*, 2351–2357.
3. Raymond, C.M.; Bryan, B.A.; MacDonald, D.H.; Cast, A.; Strathearn, S.; Grandgirard, A.; Kalivas, T. Mapping community values for natural capital and ecosystem services. *Ecological economics* **2009**, *68*, 1301–1315.
4. He, Y.; Jia, K.; Wei, Z. Improvements in Forest Segmentation Accuracy Using a New Deep Learning Architecture and Data Augmentation Technique. *Remote Sensing* **2023**, *15*, 2412.
5. Körting, T.S.; Fonseca, L.M.G.; Câmara, G. GeoDMA—Geographic data mining analyst. *Computers & Geosciences* **2013**, *57*, 133–145.

6. Khryashchev, V.; Pavlov, V.; Ostrovskaya, A.; Larionov, R. Forest areas segmentation on aerial images by deep learning. In Proceedings of the 2019 IEEE East-West Design & Test Symposium (EWDTS). IEEE, 2019, pp. 1–5.
7. Maji, S.; Malik, J. Object detection using a max-margin hough transform. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009, pp. 1038–1045.
8. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
9. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, 2015, pp. 234–241.
10. Filatov, D.; Yar, G.N.A.H. Forest and Water Bodies Segmentation Through Satellite Images Using U-Net. *arXiv preprint arXiv:2207.11222* **2022**.
11. Safarov, F.; Temurbek, K.; Jamoljon, D.; Temur, O.; Chedjou, J.C.; Abdusalomov, A.B.; Cho, Y.I. Improved Agricultural Field Segmentation in Satellite Imagery Using TL-ResUNet Architecture. *Sensors* **2022**, *22*, 9784.
12. Nichols, K.; Hosein, P. Estimating Deforestation using Machine Learning Algorithms. In Proceedings of the 2021 Second International Conference on Intelligent Data Science Technologies and Applications (IDSTA). IEEE, 2021, pp. 82–87.
13. Smith, A. Image segmentation scale parameter optimization and land cover classification using the Random Forest algorithm. *Journal of Spatial Science* **2010**, *55*, 69–79.
14. Vorotyntsev, P.; Gordienko, Y.; Alienin, O.; Rokovyi, O.; Stirenko, S. Satellite image segmentation using deep learning for deforestation detection. In Proceedings of the 2021 IEEE 3rd Ukraine Conference on Electrical and Computer Engineering (UKRCON). IEEE, 2021, pp. 226–231.
15. Adams, J.; Qiu, Y.; Xu, Y.; Schnable, J.C. Plant segmentation by supervised machine learning methods. *The Plant Phenome Journal* **2020**, *3*, e20001.
16. Khryashchev, V.; Larionov, R. Wildfire segmentation on satellite images using deep learning. In Proceedings of the 2020 Moscow Workshop on Electronic and Networking Technologies (MWENT). IEEE, 2020, pp. 1–5.
17. Bhatnagar, S.; Gill, L.; Ghosh, B. Drone image segmentation using machine and deep learning for mapping raised bog vegetation communities. *Remote Sensing* **2020**, *12*, 2602.
18. Huang, S.; Tang, L.; Hupy, J.P.; Wang, Y.; Shao, G. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *Journal of Forestry Research* **2021**, *32*, 1–6.
19. Guérin, E.; Oechslin, K.; Wolf, C.; Martinez, B. Satellite image semantic segmentation. *arXiv preprint arXiv:2110.05812* **2021**.
20. Guan, Z.; Miao, X.; Mu, Y.; Sun, Q.; Ye, Q.; Gao, D. Forest fire segmentation from Aerial Imagery data Using an improved instance segmentation model. *Remote Sensing* **2022**, *14*, 3159.
21. Sai, S.; Mikhailov, E. Texture-based forest segmentation in satellite images. In Proceedings of the Journal of Physics: Conference Series. IOP Publishing, 2017, Vol. 803, p. 012133.
22. Cheng, K.; Cheng, X.; Wang, Y.; Bi, H.; Benfield, M.C. Enhanced convolutional neural network for plankton identification and enumeration. *PLoS One* **2019**, *14*, e0219570.
23. Qassim, H.; Verma, A.; Feinzimer, D. Compressed residual-VGG16 CNN model for big data places image recognition. In Proceedings of the 2018 IEEE 8th annual computing and communication workshop and conference (CCWC). IEEE, 2018, pp. 169–175.
24. Zan, X.; Zhang, X.; Xing, Z.; Liu, W.; Zhang, X.; Su, W.; Liu, Z.; Zhao, Y.; Li, S. Automatic detection of maize tassels from UAV images by combining random forest classifier and VGG16. *Remote Sensing* **2020**, *12*, 3049.
25. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* **2014**.
26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
27. Alsabhan, W.; Alotaiby, T.; et al. Automatic building extraction on satellite images using Unet and ResNet50. *Computational Intelligence and Neuroscience* **2022**, *2022*.
28. Vapnik, V. Thc Nature of Statistical Learning Theory. Spring-Verlag, New York, 1995.
29. Cortes, C.; Vapnik, V. Support-vector networks. *Machine learning* **1995**, *20*, 273–297.

30. Joachims, T. Text categorization with support vector machines: Learning with many relevant features. In Proceedings of the European conference on machine learning. Springer, 1998, pp. 137–142.
31. Zhang, Z. Introduction to machine learning: k-nearest neighbors. *Annals of translational medicine* **2016**, *4*.
32. Mia, S.; Rahman, M.M. An efficient image segmentation method based on linear discriminant analysis and K-means algorithm with automatically splitting and merging clusters. *International Journal of Imaging and Robotics* **2018**, *18*, 62–72.
33. Anand, M.V.; KiranBala, B.; Srividhya, S.; Younus, M.; Rahman, M.H.; et al. Gaussian Naïve Bayes Algorithm: A Reliable Technique Involved in the Assortment of the Segregation in Cancer. *Mobile Information Systems* **2022**, *2022*.
34. Webb, G.I.; Keogh, E.; Miikkulainen, R. Naïve Bayes. *Encyclopedia of machine learning* **2010**, *15*, 713–714.
35. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data* **2021**, *8*, 1–74.
36. Quadeer, S. Forest Image Segmentation Forest Aerial Images for Segmentation. https://www.kaggle.com/forest-segmentation, 2021.
37. Biswas, S.; Chatterjee, S.; Majee, A.; Sen, S.; Schwenker, F.; Sarkar, R. Prediction of COVID-19 from chest CT images using an ensemble of deep learning models. *Applied Sciences* **2021**, *11*, 7004.
38. Bakasa, W.; Viriri, S. VGG16 Feature Extractor with Extreme Gradient Boost Classifier for Pancreas Cancer Prediction. *Journal of Imaging* **2023**, *9*, 138.
39. Shahana, K.; Ghosh, S.; Jeganathan, C. A survey of particle swarm optimization and random forest based land cover classification. In Proceedings of the 2016 International Conference on Computing, Communication and Automation (ICCCA). IEEE, 2016, pp. 241–245.
40. Akar, Ö.; Güngör, O. Classification of multispectral images using Random Forest algorithm. *Journal of Geodesy and Geoinformation* **2012**, *1*, 105–112.
41. Ru, F.X.; Zulkifley, M.A.; Abdani, S.R.; Spraggon, M. Forest Segmentation with Spatial Pyramid Pooling Modules: A Surveillance System Based on Satellite Images. *Forests* **2023**, *14*, 405.
42. Umar, M.; Babu Saheer, L.; Zarrin, J. Forest terrain identification using semantic segmentation on uav images **2021**.
43. Wang, J.; Fan, X.; Yang, X.; Tjahjadi, T.; Wang, Y. Semi-Supervised Learning for Forest Fire Segmentation Using UAV Imagery. *Forests* **2022**, *13*, 1573.
44. Filatov, D.; Yar, G.N.A.H. Forest and Water Bodies Segmentation Through Satellite Images Using U-Net. *arXiv preprint arXiv:2207.11222* **2022**.
45. Pyo, J.; Han, K.j.; Cho, Y.; Kim, D.; Jin, D. Generalization of U-Net Semantic Segmentation for Forest Change Detection in South Korea Using Airborne Imagery. *Forests* **2022**, *13*, 2170.
46. Shi, L.; Wang, G.; Mo, L.; Yi, X.; Wu, X.; Wu, P. Automatic Segmentation of Standing Trees from Forest Images Based on Deep Learning. *Sensors* **2022**, *22*, 6663.
47. Wang, J.; Zhu, L.; Wu, B.; Ryspayev, A. Forestry Canopy Image Segmentation Based on Improved Tuna Swarm Optimization. *Forests* **2022**, *13*, 1746.

3.2.2 Conclusion

This study employs a hybrid methodology that combines deep learning models with conventional machine learning classifiersfor the purpose of identifying forested and non-forested regions within an aerial satellite image sourced from the Deep Globe challenge dataset. The researchers employed a combined methodology involving VGG16 and ResNet50 deep learning models to extract a comprehensive set of features. These features were then utilised by machine learning classifiers to effectively segment an aerial satellite image into distinct forest and non-forest regions. The performance of the models was evaluated using metrics such as Intersection over Union (IoU), accuracy, Root Mean Square Error (RMSE), and Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) curves. The model utilising Random Forest (RF) emerged as the superior performer, attaining an accuracy rate of 94%, Intersection over Union (IoU) score of 91%, Root Mean Square (RMSE) value of 0.245, and ROC_AUC value of 0.98. The model's great efficacy suggests its potential for detecting smoke, veld fires, and conducting water segmentation.

4 Classification of Satellite Forest Images

This Chapter presents published works in the classification of forest images into their respective categories.

4.1 Forest Image Classification based on Deep learning and XGBoost Algorithm

4.1.1 Introduction

This paper looked at the classification approach suitable for solving forest classification problems. The aim of the paper is to design a hybrid model that harmonizes both deep learning and machine learning approaches with the goal of increasing forest image classification accuracy. The model was optimized to improve its performance.

Forest Image Classification Based on Deep Learning and XGBoost Algorithm

Clopas Kwenda( ), Mandlenkosi Victor Gwetu, and Jean Vincent Fonou-Dombeu

School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa
221072651@stu.ukzn.ac.za, {gwetum,fonoudombeuj}@ukzn.ac.za

Abstract. Deep learning and machine learning methods have been recently used in forest classification problems, and have shown significant improvement in terms of efficacy. However, as attributed from the literature, they have the challenge of having insufficient model variance and restricted generalization capabilities. The goal of this study is to improve the accuracy of forest image classification through the development of a hybrid model that incorporates both deep learning and machine learning techniques. This study has proposed an ensemble approach of the Deep Learning technique (ResNet50 in particular), and machine learning model (specifically XGBoost) to increase the prediction capability of classifying satellite forest images. The sole purpose of ResNet50 is to generate a set of features that will in turn be used by the XGBoost algorithm to perform the classification process. The XGBoost algorithm was compared against a fully connected ResNet50 model and other classifiers such as random forest (RF) and light gradient boost machine (LGBM). The best classification results were obtained from XGBoost (0.77), followed by RF (0.74), LGBM (0.73), and ResNet50 (0.59).

Keywords: Machine learning · feature extraction · Convolutional Neural Networks · Image Processing

1 Introduction

Forests remain a key natural resource for both developing and developed countries as their wood and forestry products contribute significantly towards a country's Gross National Product (GDP). Both satellite and aerial images play a pivotal role when it comes to monitoring and evaluation of forests and other vegetation. Such images have made huge significant progress in solving remote sensing science classification problems. Data obtained from features such as spectral, radiometric, and spatial is usually used to perform the forest classification process [1]. Image classification refers to the process of labeling each image into its corresponding category or class [2]. Image segmentation is centered on pixel level classification, whereas image classification involves classifying the entire object into one of the given classes. In general, the majority of classification methods employ the technique of assessing and evaluating the image's content

J. Mikyška et al. (Eds.): ICCS 2023, LNCS 14076, pp. 217–229, 2023.
https://doi.org/10.1007/978-3-031-36027-5_16

and then marshaling pixels into their respective categories. The new instance is classified based on an already trained data set whose classes are known. In general, an image is classified into only one of the predefined classes; however, in some cases, an image can be classified into multiple classes, which are referred to as multi-label classes [2]. In spite of the existence of many algorithms used in the classification of vegetation images, there are limited studies that have employed the ensemble machine learning approach in the classification of satellite forest images. Therefore, the purpose of this study is to report findings obtained from an ensemble approach of XGBoost algorithm and ResNet50 technique for the classification of satellite forest images. The new ensemble classifier approach's performance is evaluated against other classifiers such as Random Forest (RF) and Light Gradient Boost Machine (LGBM) in terms of classification accuracy. Different classes (bare-land, logged forest, shrubs, woodlands, and degraded forest) have been identified, and the ensemble learning approach for satellite forest image classification has been assessed by estimating image classification accuracy for different class labels. The rest of the paper is structured as follows. Section 2 deals with related work. Section 3 describes the flow of the proposed study. Section 4 describes the overview of the model architecture. Results and Discussion are presented in Sect. 5. Section 6 concludes the paper.

2 Related Studies

A study by [3] adopted the Random Forest (RF) algorithm to perform image classification on multi-spectral images obtained from Ikonos and QuickBoard data sets. The algorithm's performance was evaluated against results obtained from Gentle AdaBoost (GAB), Maximum Likelihood Classification (MLC), and Support Vector Machine (SVM) algorithms, and RF gave the best result compared to others. The major issue arising in their study was feature extraction. Features were generated using the Random Feature Selection technique. The main limitation of such a technique is giving equal or similar importance to correlated features. To solve this problem, the proposed study has adopted ResNet50 deep learning technique which excels at producing apt and specific features required to solve image classification problems.

[4] employed a deep learning supervised approach on Unmanned Aerial Vehicle (UAV) satellite images for forest area classification. The deep learning stacked Auto-encoder showed significant potential with regard to forest area classification accuracy. However, the major limitation of the deep learning model is that it requires high computational facilities as compared to machine learning algorithms. As a way of solving this challenge, this study is designed in such a way that the image classification process which is the major task that requires high computational capabilities is performed by the XGBoost machine learning algorithm, while the feature extraction part is performed by the ResNet50.

[5] developed a deep learning model for image classification of VHR (very high resolution) images obtained using UAV. The study was against the backdrop that UAV data sets have been found to be very useful for forest feature identification

attributed to their high spatial resolution. Pre-processed data sets of forests of Nagli area were used for the study. The deep learning model incorporated a stacked Auto-encoder to perform image classification. Results showed that the deep learning technique outperformed other machine learning algorithms in terms of accuracy. Through Cross Validation the deep learning model achieved an accuracy 97%. The study's limitation was that it included all features for classification rather than only appropriate features, resulting in an overhead in terms of the model's time complexity. To address this problem, this study adopted the ResNet50 model to generate a set of features required for the forest image classification problem. The learning process of this model is such that the upper first layers are designed to learn general features and the last lower layers are designed to learn specific features. The final feature vector obtained from the ResNet50 model is specifically related to solving a specific classification problem.

When applied to image classification, traditional artificial neural networks, and machine learning approaches face difficulties in processing massive images for feature extraction, resulting in low efficiency and classification accuracy [6]. [6] proposed a deep learning model for image classification with the goal of providing support for classifying large image datasets. The study discussed various types of convolutional neural networks and their applications in image processing. The model was refined by adjusting parameters for feature extraction and by undergoing a process of noise reduction. This study optimized the proposed deep learning model in order to improve the model's classification efficiency and accuracy. The proposed model outperformed other models such as AleXNet and LeNet in terms of classification accuracy. Classification accuracy was also assessed before and after the optimization of the deep learning model. The results revealed that the optimized model significantly improved image classification accuracy. However, the model had challenges in classifying dynamic targets in a complex environment.

Convolutional neural networks adopted under transfer learning usually compress high-resolution input images [7]. A downsampling operation like this usually results in information loss, which affects image classification accuracy. [7] proposed a CNN model based on wavelets domain inputs to solve this problem. During the image pre-processing stage, the wave packet transform was used to extract information from input images. Some subband image channels were chosen as inputs for conventional CNNs with the first several convolutional layers removed, allowing the networks to learn directly in the wavelet domain. The model achieved a classification improvement of 2.15% and 10.26%, respectively on Caltech-256 dataset and Describable Textures Dataset. However, the model suffered huge problems in terms of training costs due to wavelet transform operations that were applied to each image generated through the augmentation process. To address this issue in the proposed study, output images were obtained from the third batch normalization layer of the ResNet50 architecture, where an image would not have been significantly compressed. [8] used an object-based random forest algorithm to identify eight forest types from freely available remote sensing images in Wuhan, China. The images were obtained using Sentinel-1A,

Sentinel-2A, and Landsat 8 sensors. Results obtained indicated that a single sensor cannot obtain satisfactory results. Phenological and topographic information were used in the hierarchical classification to improve discrimination between different forest types. The final forest-type map was obtained using a hierarchical strategy and had an overall accuracy of 82.78%. However, the model encountered the issue of misclassification on types with similar spectral characteristics. This issue is attributed to the study's use of only the NDVI as the primary feature indicator for image classification. This challenge again is addressed in this study by adopting the ResNet50 model.

3 Proposed Model

The study proposes a hybrid machine learning technique for forest-type image classification that combines convolutional deep learning, specifically ResNet50, and traditional machine learning (XGBoost). Convolutional neural networks are widely used in generating features for solving specific classification problems [9]. Therefore in the same vein, the ResNet50 model was adopted in this study to generate a set of features for the XGBoost to perform the image classification task. The XGBoost algorithm was adopted only to perform the image classification process task. Traditional machine learning algorithms outperform deep learning techniques in terms of classification accuracy for a limited data set. Hence the study adopted the XGBoost (machine learning algorithm) to perform the classification task. Because the study uses limited forest images, the basic idea of the model is that CNN produces a feature vector, and then the XGBoost performs the image classification process. Traditional machine learning algorithms used in image classification include Support Vector Machines (SVM), decision trees (DT), extreme gradient boost (XGBoost), random forest (RF), and k-nearest neighbor (KNN). [10] conducted a study to compare the efficacy and effectiveness of LGBM and XGBoost in remote sensing image classification to RF, KNN, and SVM. Efficacy levels of XGBoost and LGBM were above 90%, while the other algorithms had efficacy levels below 90%. It is against this backdrop that the proposed model has advocated towards XGBoost. The ensemble model was used to perform multi-label image classification on forest images from the categories of logged forest, bare land, degraded forest, woodlands, shrubs, and grassland. The proposed algorithm is shown in Fig. 1. The proposed ensemble learning approach for multi-label image classification has the following key features.

- ResNet50 is adopted under the transfer learning technique
- ResNet50 is used for feature extraction and XGBoost is used to perform the classification task.

3.1 Multi-label Image Classification

Multi-label classification is when a test forest image is assigned to a correct category from a set of categories. Fine-tuning done to the model to enable the

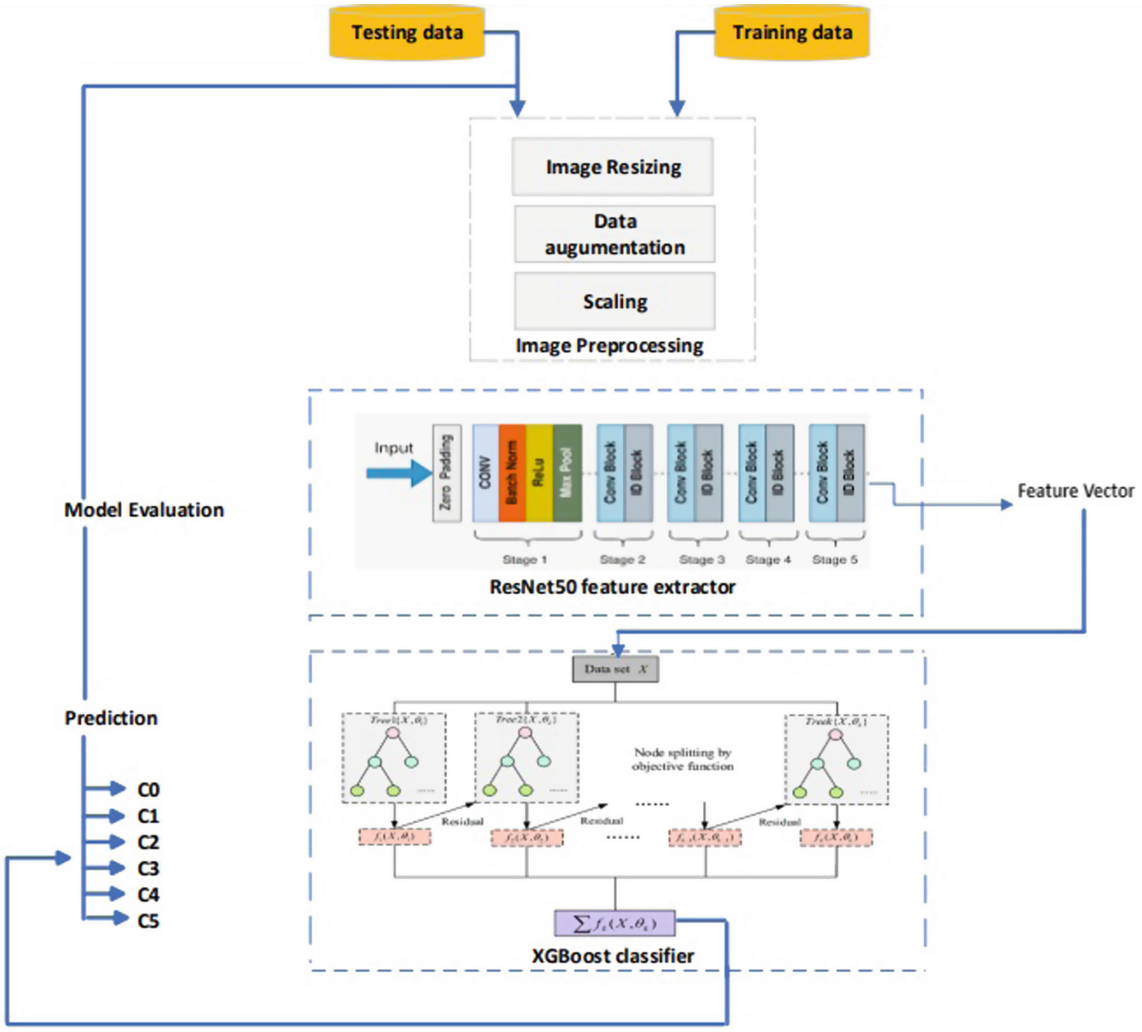


Fig. 1. Proposed ensemble hybrid algorithm for forest image classification.

classification processes involves converting class label strings to integer discrete values. Such conversion is made possible by applying the transform LabelEncoder function adopted from sklearn in Python on the class label vector set. The inverse transform function was invoked in the prediction phase for visualization purpose. Figure 2 shows a sample of forest-type image data set that was used in the study.

3.2 Pre-processing

Since there is no publicly available forest image data set [11], different types of forest images were obtained from the internet. All images were resized to 256 X 256 pixels since the images were of different sizes. Class labels were used as categorical data in this study, and the label encoder technique was used to convert non-numeric categorical data to numeric values. The class labels were transformed into a vector of values 0 through 5. Most machine learning algorithms require labels to numerical integer values. Table 1 represents the labeled classes. Scaling features in machine learning is one of the most critical steps in

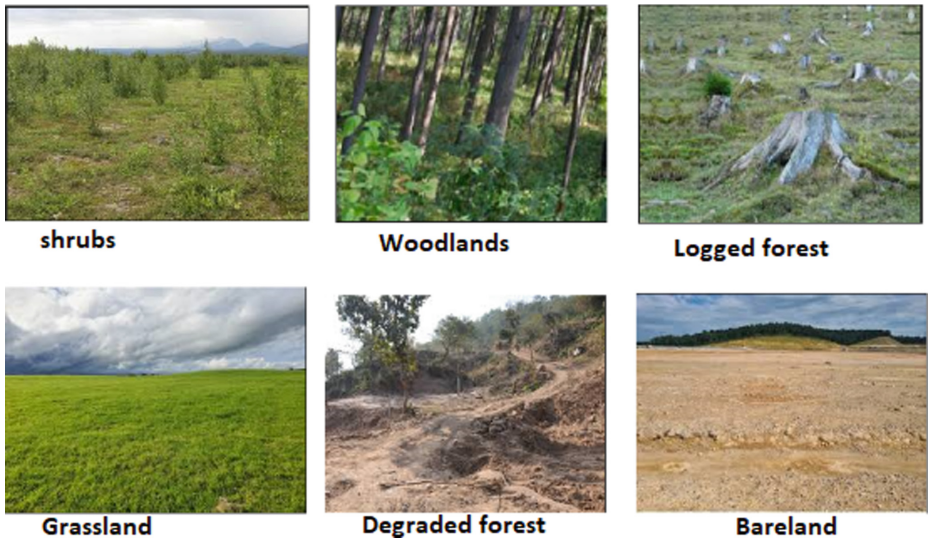


Fig. 2. Sample of forest type image dataset.

pre-processing of data as most models are sensitive to the magnitude of features. Scaling refers to bringing all values to a uniform scale. All images were scaled by dividing image pixel values by 255 since the images were 8-bit images such that the scaling was in the range between 0 and 1. Data augmentation is a process of generating more image data sets from already existing images. 30 images for each category were downloaded from the internet and 90 more images respectively were generated through the data augmentation process with settings prescribed in Table 2. The forest image data set was split in a way that 80% was reserved for training and 20% for testing.

Table 1. Labels of forest type images

Value	Class
0	bareland
1	degraded forest
2	grassland
3	woodlands
4	logged forest
5	shrubs

4 Overview of the Model Architecture

This section provides a description of algorithms that were harmonized together to form the proposed hybrid model.

Table 2. Data Augmentation Properties

Property	Value
rotation_ range	45
width_shift_range	0.2
height_shift_range	0.2
zoom_range	0.2
horizontal_flip	True
fill_mode	reflect

4.1 The XGBOOST Algorithm

XGBoost has become predominant in the fraternity of machine learning. It is highly preferred as an alternative to Light Gradient Boost Machines (LGBMs) due to its high execution speed and performance. During the CPU's running time, the XGBoost algorithm employs a parallel computing technique for subsequent tree construction. It uses the 'maxdepth' criteria, instead of the traditional stopping criterion first, and the tree pruning process is initiated from a backward direction. Such a technique significantly improves the computational speed of XGBoost over other LGBM frameworks. Another strength of XGBoost is that it uses the training loss function to automatically learn the best missing values, hence it has the ability to handle different sparsity patterns in the data provided as input efficiently. The XGBoost algorithm uses the following equations for classification:

$$x(t) \approx x(s) + x'(s)(t-a) + \frac{1}{2}x^n(s)(t-s)^2, \tag{1}$$

$$\zeta \simeq \sum_{i=1}^{n}[l(q_i, q^{t-1}) + r_i x_t(t_i) + \frac{1}{2}s_i x_t^2(m_i)] + \omega(x_t + C), \tag{2}$$

where C is constant, m_i is the input, Ω(x) is the complexity of the tree. r_i and s_i are defined as follows:

$$r_i = \delta\hat{z}_i^{(b-1)}.\int(z_i\hat{z}_i^{n(b-1)}, \tag{3}$$

$$s_i = \delta\hat{z}_i^{(b-1)}.\int(z_i\hat{z}_i^{n(b-1)}, \tag{4}$$

where z_i represents the real value obtained from the training data set. [12] conducted a comparative performance assessment of the XGBoost algorithm, random forest, logistic regression, and standard gradient boosting, and the XGBoost algorithm was found to be most efficient against all other algorithms. It is against this backdrop that the study has settled for the XGBoost technique.

4.2 ResNet50 Network Architecture

A CNN composed of 50 layers is referred to as ResNet-50. Such a deep network with so many layers suffer from network degradation problem. The network is made up of stalked residual blocks. It performs its function with identity shortcut connections that jump one or more layers during the training phase using the residual connections. Intermediate layers have the learning ability to self-adjust their weights to values closer to zero such that the residual block becomes an identity function. The residual skip connections in the ResNet50 architecture helps solves the problem of vanishing gradient experienced in deep neural networks. It is against this backdrop that the study has adopted the ResNet50 model in the framework. Due to the limited labeled training data set, the study sped up the learning process by adopting the ResNet50 under the transfer learning technique pre-trained on the ImageNet database. ResNet50 architecture is widely known for producing good features for solving classification problems. Features are produced in a way that the upper learns lower-level features and lower layers learn specific features.

5 Metrics for the Study

The multi-class classification evaluation metrics used are accuracy, precision, recall, F1-score, mean average error, root mean square, and confusion matrix. Mean Absolute Error is a measure of the difference between the predicted value and the actual value. It gives an error associated with a predicted image

The root mean square error (RMSE) and the mean absolute error (MAE)are two widely standard metrics used to assess the performance of a model. MEA gives an error associated with a predicted image while RMSE as the name suggests gives the mean square of all errors. Considering a set of m observations $x(x_i, i = 1, 2, 3...m)$ and the corresponding model predictions $\hat{x}$ the MAE and RMSE are

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|x_i - \hat{x}_i| \tag{5}$$

$$RMSE = \frac{1}{m}\sqrt{\sum_{i=1}^{m}(x_i - \hat{x_i})^2} \tag{6}$$

Precision being closely related to the measure of quality and recall to the measure of quantity, these two metrics are expressed as follows:

$$precision = \frac{TP}{TP + FP} \tag{7}$$

$$recall = \frac{TP}{TP + FN} \tag{8}$$

where TP is true positives, FP denotes false positives and FP denotes false negatives. F1-Score calculates the harmonic average between recall and precision rates and is expressed as follows:

$$F1 - Score = 2 * \frac{precision * recall}{precision + recall} \tag{9}$$

Accuracy is the overall measure of the model performance and it is expressed as:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

Table 3. Metrics for Random Forest

Label	Precision	Recall	F1-Score
0	0.88	0.54	0.67
1	0.50	0.55	0.52
2	0.67	0.80	0.73
3	0.78	0.70	0.74
4	0.71	0.85	0.77
5	0.83	1.00	0.88

6 Results and Discussion

Discussion and results obtained by the study are presented in this section. Table 3 shows that the RF algorithm returns the highest precision for category 0 and subsequently followed by category 5, 3, and 4 respectively, and performed poorly for category 1. Precision is also referred to as the measure of quality. Most of the images in Category 1 were misclassified into Category 0 and this is most likely due to image ambiguity between bare land and degraded forests. However, for recall, category 5 received the most relevant images followed by categories 4, 2, and 3 respectively. Recall is also referred to as the measure of quantity. F1-score provides a balance between precision and recall in relation to positive classes. RF achieved the highest F1-Score for category 5, followed by categories 4, 3, 2, and 1 respectively. Table 4 shows that XGBoost obtained high precision for category 0 with 0.86, i.e. slightly lower than RF. Categories 2, 3, and 4 obtained good quality results in terms of precision as all the scores are above 0.7. Similar to the RF algorithm, the XGBoost algorithm obtained poor results for category 1, and the same reason attributed to poor results in category 1 in RF is also attributed here. The general performance of XGBoost in terms of recall and F1-score is generally the same as with RF. Table 5 shows that LGBM performed poorly for category 1 in terms of precision, recall, and F1- score. That is, the algorithms

Table 4. Metrics for XGBoost

Label	Precision	Recall	F1-Score
0	0.86	0.69	0.77
1	0.50	0.64	0.56
2	0.82	0.70	0.76
3	0.82	0.70	0.76
4	0.75	0.90	0.82
5	0.79	0.95	0.86

Table 5. Metrics for LGBM

Label	Precision	Recall	F1-Score
0	0.75	0.69	0.72
1	0.33	0.18	0.24
2	0.78	0.70	0.74
3	0.82	0.70	0.76
4	0.63	0.85	0.72
5	0.80	1.00	0.89

failed to distinguish clearly between bare land and degraded forests. For the remaining categories, the algorithm obtained good promising results because on average the values obtained were above 70% for all the metrics. The performance of a fully linked ResNet50 was subpar in comparison to that of other classifiers. As presented in Table 6, the model performed the poorest in Category 1 as it registered a zero for all the metrics that were considered.

Table 6. Metrics for ResNet50

Label	Precision	Recall	F1-Score
0	0.37	0.38	0.38
1	0.00	0.00	0.00
2	0.44	0.60	0.51
3	0.75	0.75	0.75
4	0.76	0.65	0.70
5	0.76	0.95	0.84

Accuracy, MAE, and RSME are the most commonly used metrics to evaluate the performance of the model. The hybrid model with XGBoost outperformed the other algorithms in terms of Accuracy, MAE, and RMSE, which obtained values of 0.77, 0.56, and 1.30, respectively as presented in Table 7. Our proposed model's accuracy outperformed the model proposed by [13]. The model

Table 7. Metrics for Classifiers

Classifier	MAE	RMSE	Accuracy
Random Forest	0.63	1.86	0.74
XGBoost	0.56	1.30	0.77
LGBM	0.67	0.40	0.73
ResNet50	0.97	1.68	0.59

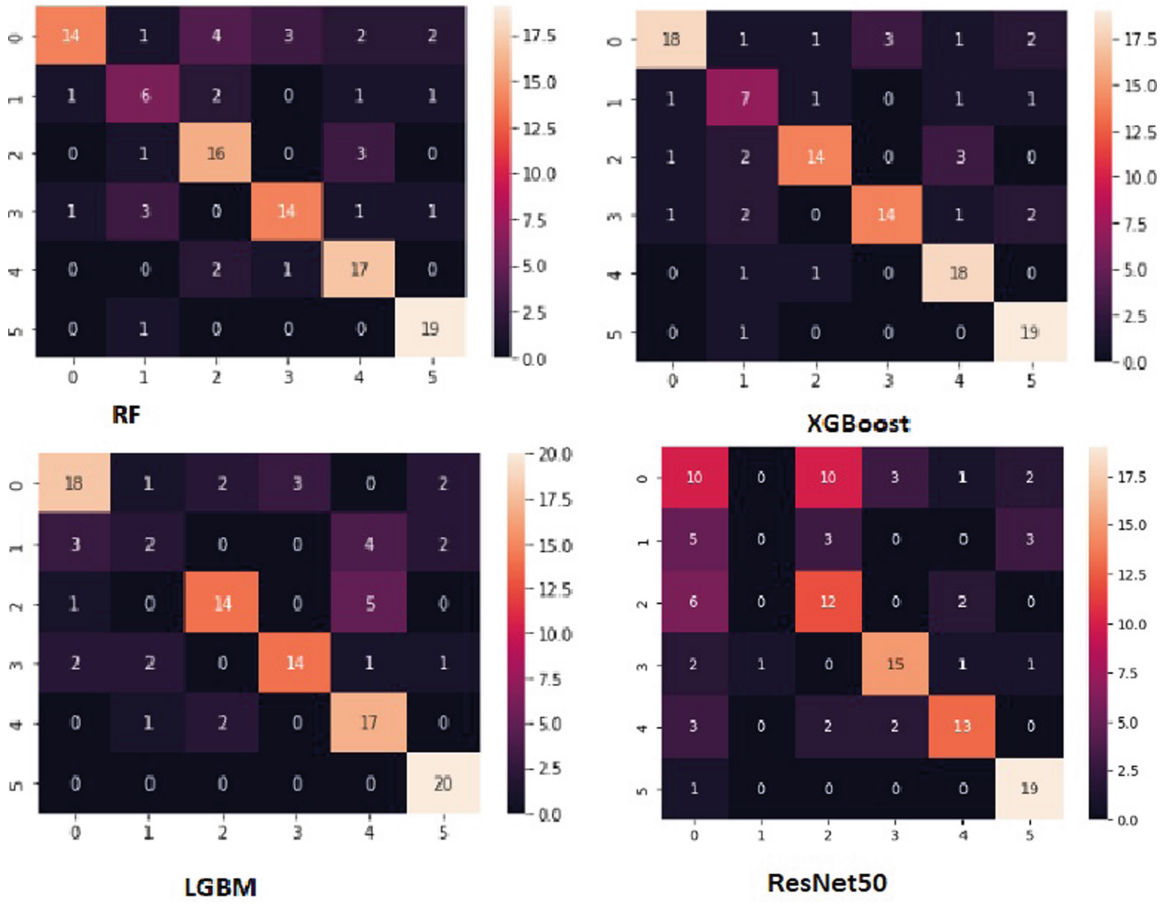


Fig. 3. Confusion Matrix results obtained from RF, XGBoost, LGBM, and ResNet50.

used CNN and Multitemporal High-Resolution Remote Sensing Images to classify individual Tree Species and it obtained an overall accuracy of 75.1% for seven tree species using only the WorldView-3 image data set. The classification accuracy of our proposed model also performed better compared to the results obtained by [3]. Their study achieved a classification accuracy of 68% for classifying multispectral images using Support Vector Machine (SVM). However, the ResNet50 deep learning model proposed by [11] for classifying forest image data set outperformed our model as it achieved an accuracy of 92% for classifying forest images belonging to 3 categories. Such high accuracy could be attributed to the fact that the model was applied on only 3 categories whilst our model was applied to 6 different categories. The performance of a classification algo-

rithm is reflected in a two-dimensional table called the confusion matrix. It is important for summarizing and visualizing a classification algorithm's results. The confusion matrix results as presented in Fig. 3 show that there was high misclassification for category 1 by all the algorithms, with the worst performance by ResNet50. Apart from Category 1, the performance of LGBM and XGBoost in the other categories is generally the same.

7 Conclusion

An ensemble learning approach of ResNet50 and XGBoost was developed to classify forest images into their respective categories. ReNet50 adopted under the transfer learning technique was used as a feature generator, while the XGBoost algorithm was used to perform the forest image classification process. The model was evaluated against a fully connected ResNet50 and other baseline classifiers such as LGBM and Random Forest. The proposed ensemble learning technique achieved a classification accuracy of 77%. Therefore the proposed model in this study can be used to classify forest images since it recorded high classification accuracy and low RMSE and MAE values as compared to other classifiers. For future studies, it is recommended to incorporate an ensemble stack of CNNs for generating plausible features for subsequent image classification. This approach would significantly increase the scope of features required to perform the image classification process.

References

1. Drobnjak, S., Stojanović, M., Djordjević, D., Bakrač, S., Jovanović, J., Djordjević, A.: Testing a new ensemble vegetation classification method based on deep learning and machine learning methods using aerial photogrammetric images. Front. Environ. Sci. **702** (2022)
2. Rout, A.R., Bagal, S.B.: Natural scene classification using deep learning. In: 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), pp. 1–5. IEEE (2017)
3. Akar, Ö., Güngör, O.: Classification of multispectral images using random forest algorithm. J. Geodesy Geoinf. Sci. **1**(2), 105–112 (2012)
4. Haq, M.A., Rahaman, G., Baral, P., Ghosh, A.: Deep learning based supervised image classification using UAV images for forest areas classification. J. Indian Soc. Remote Sens. **49**(3), 601–606 (2021)
5. Zhang, X., Chen, G., Wang, W., Wang, Q., Dai, F.: Object-based land-cover supervised classification for very-high-resolution UAV images using stacked denoising autoencoders. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **10**(7), 3373–3385 (2017)
6. Lv, Q., Zhang, S., Wang, Y.: Deep learning model of image classification using machine learning. Adv. Multimed. **2022** (2022)
7. Wang, L., Sun, Y.: Image classification using convolutional neural network with wavelet domain inputs. IET Image Process. **16**(8), 2037–2048 (2022)

8. Liu, Y., Gong, W., Hu, X., Gong, J.: Forest type identification with random forest using Sentinel-1A, Sentinel-2A, multi-temporal Landsat-8 and DEM data. Remote Sens. **10**(6), 946 (2018)
9. Wang, P., Fan, E., Wang, P.: Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. Pattern Recogn. Lett. **141**, 61–67 (2021)
10. Łoś, H., et al.: Evaluation of XGBoost and LGBM performance in tree species classification with sentinel-2 data. In: 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 5803–5806. IEEE (2021)
11. Tang, Y., Feng, H., Chen, J., Chen, Y.: ForestResNet: a deep learning algorithm for forest image classification. J. Phys: Conf. Ser. **2024**(1), 012053 (2021). IOP Publishing
12. Morde, V.: XGBoost algorithm: long may she reign! (1999)
13. Guo, X., Li, H., Jing, L., Wang, P.: Individual tree species classification based on convolutional neural networks and multitemporal high-resolution remote sensing images. Sensors **22**(9), 3157 (2022)

4.1.2 Conclusion

A model developed as a result of integrating ResNet50 and XGBoost algorithm managed to classify forest images into the respective categories, as it achieved a classification accuracy of 77%. ResNet50 model was used as a feature generator, while XGBoost was used to perform the classification process. In order to enhance future research, it was advisable to include an ensemble stack of Convolutional Neural Networks (CNNs) to provide credible features that may be utilized for further forest image categorization. The adoption of this strategy would lead to a notable expansion in the range of features that are necessary for performing forest image classification.

4.2 Ontology with Deep Learning for Forest Image Classification

4.2.1 Introduction

This paper describes the state-of-the-art model that incorporates ontologies and deep neural networks for the purpose of classifying forest images. The use of ontologies is against the backdrop that most forest image classification approaches neglect the concept of semantics whilst forest image categories treated as independent have a strong semantic overlap. An ensemble approach of Xception, ResNet50, and VGG16 is used to produce a set of features for the classifiers trained through ontology to perform the image classification process.

This paper has been published in the MDPI (applied Journal).

applied sciences

MDPI

Article

Ontology with Deep Learning for Forest Image Classification

Clopas Kwenda *, Mandlenkosi Gwetu † and Jean Vincent Fonou-Dombeu †

School of Mathematics, Statistics and Computer Science, University of KwaZulu Natal, Pietermaritzburg 3209, South Africa; fonoudombeuj@ukzn.ac.za (J.V.F.-D.)
* Correspondence: 221072651@stu.ukzn.ac.za; Tel.: +27-0612039734
† These authors contributed equally to this work.

Abstract: Most existing approaches to image classification neglect the concept of semantics, resulting in two major shortcomings. Firstly, categories are treated as independent even when they have a strong semantic overlap. Secondly, the features used to classify images into different categories can be the same. It has been demonstrated that the integration of ontologies and semantic relationships greatly improves image classification accuracy. In this study, a hybrid ontological bagging algorithm and an ensemble technique of convolutional neural network (CNN) models have been developed to improve forest image classification accuracy. The ontological bagging approach learns discriminative weak attributes over multiple learning instances, and the bagging concept is adopted to minimize the error propagation of the classifiers. An ensemble of ResNet50, VGG16, and Xception models is used to generate a set of features for the classifiers trained through an ontology to perform the image classification process. To the authors' best knowledge, there are no publicly available datasets for forest-type images; hence, the images used in this study were obtained from the internet. Obtained images were put into eight categories, namely: orchards, bare land, grassland, woodland, sea, buildings, shrubs, and logged forest. Each category comprised 100 images for training and 19 images for testing; thus, in total, the dataset contained 800 images for training and 152 images for testing. Our ensemble deep learning approach with an ontology model was successfully used to classify forest images into their respective categories. The classification was based on the semantic relationship between image categories. The experimental results show that our proposed model with ontology outperformed other baseline classifiers without ontology with 96% accuracy and the lowest root-mean-square error (RMSE) of 0.532 compared to 88.8%, 86.2%, 81.6%, 64.5%, and 63.8% accuracy and 1.048, 1.094, 1.530, 1.678, and 2.090 RMSE for support-vector machines, random forest, k-nearest neighbours, Gaussian naive Bayes, and decision trees, respectively.

Keywords: ontology; feature extraction; convolutional neural networks; image classification

Citation: Kwenda, C.; Gwetu, M.; Fonou-Dombeu, J.V. Ontology with Deep Learning for Forest Image Classification. *Appl. Sci.* **2023**, *13*, 5060. https://doi.org/10.3390/app13085060

Academic Editor: Yu-Dong Zhang

Received: 21 March 2023
Revised: 14 April 2023
Accepted: 15 April 2023
Published: 18 April 2023

1. Introduction

The majority of classification algorithms treat classes or categories of images independently both in terms of visual and semantic aspects [1]. In contrast, human beings use semantic relationships when classifying images into their respective categories [2]. For instance, it might seem unreasonable to distinguish "tree" from "vegetation" since a "tree" is a kind of "vegetation". Generally, human beings use features to distinguish different kinds of objects. For example, the NDVI (normalized difference vegetation index) is an essential feature for distinguishing between vegetation and water, while shapes can discriminate between broad-shaped leaves and needle-shaped leaves. Most classification algorithms achieved better performance results on easy image classification datasets such as Caltech 256 [3] and Caltech 101 [4]; however, they neglected the concept of semantics [5], which led to poor results on fine-grained images [6]. An ontology is a hierarchical structure of a particular domain that consists of all classes or categories as well as relationships such as "is-a" and "kind-of". It captures semantic relationships between classes or categories in a manner that is close to human perception.

The adoption of ontologies in image classification algorithms incorporates semantics tools, thus leading to increases in image classification accuracy. Traditional ontology-based algorithms hugely suffer from the problem of error propagation because these ontology classifications are based on having classifiers at every ontological node, such that node subcategories are discriminated. Such errors were caused by intra-class variations of super-categories. The previous uses of ontologies in classification focused on improving speed rather than accuracy. Ontologies revolve around the use of semantic relationships, and data are expressed more at the semantic level, thus accounting for better classification. This study proposes an image classification model based on ontology and an ensemble stack of the Xception, VGG16, and ResNet50 models, which are employed to generate a set of features that are used by merged classifiers driven by taxonomic relationships in an ontology to improve image classification accuracy. The three pre-existing models, Xception, VGG-16, and ResNet50, have been adopted in this study via transfer learning because they have an innately dissimilar architecture that abstracts unrelated information from the images used for the classification purposes [7]. Some potential applications of ontological bagging in forestry include species classification. Information relating to forest tree species plays a critical role in ecology and forest management [8]. Our proposed ontological bagging model can be employed to improve the accuracy of species classification in forestry. Vegetation is an important part of an ecosystem because it provides oxygen and a suitable place for human beings to live [9]. Therefore, information concerning vegetation is very critical; hence, our proposed model can be used to classify vegetation into different types and categories. Our model can also be used to classify fruits into their respective categories. Fruit classification plays an important role in many industrial applications, including supermarkets and factories. The importance of fruit classification can be seen in people with special dietary requirements; in this case, they can be assisted in selecting categories of fruits [10]. The contribution of the study is summarised as follows:

- We integrate semantic ontologies and aggregate outputs from hypernym–hyponym classifiers to increase image category distinction and also eliminate error propagation problems, hence increasing image classification accuracy.
- We propose a new approach to image classification that uses an ensemble of Xception, ResNet50, and VGG16, whereby features obtained from Xception, ResNet50, and VGG16 are integrated together to produce all possible features, which are, in turn, used by an ontological bagging algorithm for subsequent classification.

The rest of the paper is structured as follows. Section 2 discusses related work. Section 3 discusses the dataset used for the study. Section 4 describes the deep learning architectures. Section 5 describes the ontological bagging algorithm used in the study. Section 6 describes the proposed algorithm. Section 7 outlines the experimental setup. Section 8 describes the experimental results. Section 9 discusses the results obtained from the experiment. Section 10 concludes the paper.

2. Related Works

Image classification has received much attention in the fields of computer vision and image processing [11–16]. A study [11] developed a model that harmonized ontology and HMAX features to perform image classification using merged classifiers. The basic idea behind the model was to exploit ontological relationships that exist between image classes or categories. For better discrimination between classes, the procedure involved training visual feature classifiers and merging outputs of hypernym–hyponym classifiers. The model included three components: (1) feature extraction, (2) ontology building, and (3) image classification. The visual features were obtained from the training dataset, and ontology building was carried out by mainly following the process of concept extraction and relationship generation. Visual features extracted from the training set and the ontology were used to perform image classification using a linear orange support-vector machine (SVM) classifier. In terms of accuracy, the model achieved an accuracy of 0.63, while the baseline method without ontology obtained an accuracy of 0.59. However, as coined by [12],

HMAX does not perform very well in terms of feature extraction over a limited dataset. To circumvent this shortcoming, the proposed model in this study has adopted an ensemble of CNNs to generate features for subsequent image classification.

Another study [1] proposed an ontological random forest algorithm for forest image classification. The algorithm's basic idea was that the semantic relationships between categories determined the splitting of the decision tree. Multiple-instance learning was used to provide a learning platform for generating hierarchical features that were then used to capture visual dissimilarities at various concept levels. Semantic splitting was used to build decision trees, and semantic relationships were used to learn hierarchical weak features. The experimental results showed that the approach not only outperformed state-of-the-art approaches but was also capable of identifying semantic features at different concept levels. The drawback of this study was that feature generation was hugely dependent on weak attribute learning. To solve this problem, the proposed study used an ensemble deep learning approach to generate all plausible features for subsequent image classification.

An algorithm that automatically builds image classification trees was proposed in [17]. A set of categories was recursively divided into two minimally confused subsets and achieved 5–20-fold speedups over other methods. Other authors [18] used lexical semantic networks to integrate knowledge about inter-category relationships into the learning process of visual appearance. A semantic hierarchy of discriminative classifiers was used for object detection. The challenge encountered was that object recognition was marred by the fact that the algorithm did not support weak attribute reasoning. To overcome this challenge, the proposed study incorporated the bagging algorithm because it has the ability to learn weak attributes.

A new formalism that incorporated hierarchy and exclusion (HEX) graphs to perform object classification by exploiting the rich structure of real-world labels was introduced [19]. The new formalism has the ability to capture semantic relationships between any two labels on the same object. Results obtained from the model showed an improvement in object classification as a result of exploiting label relationships. However, the major limitation of the approach is that it is too general in nature and is only limited to domains with hierarchical and exclusion relationships.

A study [20] developed a deep learning model for multiple-instance learning (MIL) frameworks, whose goal was to perform vision tasks such as classification and image annotation. In the model, each image object uses two instance sets of object proposals and text annotations to perform vision tasks. The main merit of the model is its ability to learn relationships between objects and annotation proposals. The study contributed extensively to solving computer vision tasks, and it performed well both in image classification and image auto-annotation. However, the shortcomings of the model were that it required fine tuning on the orange dataset, which is time-consuming.

A unified CNN-RNN model for multi-class image classification was proposed in [21]. The classification process consisted of learning semantic redundancy and the co-occurrence dependency in an end-to-end way. The model has the ability to obtain semantic level dependency and image label relevance by learning the joint image label embedding. The model could also be trained from scratch to integrate both pieces of information in a unified framework. The results obtained show better performance in terms of classification than the state-of-the-art multi-label classification model.The shortcomings of the model were that it fails to make a prediction on small objects that have little covariance dependencies with other, larger objects.

Considering that microscopic imaging technology is rapidly advancing, bio-image-based approaches to protein subcellular localization have sparked a lot of interest. However, there are fewer techniques for predicting protein location, with the majority of them relying on automatic single-label classification. Therefore, a study [22] developed an artificial intelligence (AI)-based stacked ensemble approach for the prediction of protein subcellular localization in confocal microscopy images. The ensemble approach was built by stacking ResNet152, DenseNet169, and VGG16 as base learners, and their predictions were inte-

grated and fed as input to the meta-learner. The model was implemented on an image dataset obtained from Human Protein Atlas Image Classification on Kaggle and attained precision, F1-score, and recall of 0.71, 0.72, and 0.70, respectively. The main difficulty encountered in the study was a huge imbalance of images in the image categories, as some classes had very few images, which were insufficient to train the model. In our study, we used a data augmentation technique to determine image balance across categories. The evolution of AI applications has significantly increased the utilization of smart imaging devices. Convolutional neural networks (CNNs) are widely used in image classification because they do not require any handcrafted features to influence performance. However, fruit classification in the horticulture field suffers from the significant disadvantage of requiring an expert with extensive knowledge and experience. To address this issue, a study [23] developed MobileNetV2 with a deep learning technique for fruit image classification. The study did not require the intervention of experts. The model used 26,149 images of 40 different fruit types from a Kaggle public dataset and achieved 99% accuracy. The model could be improved using a larger variety of fruits for broader fruit classification.

The idea of annotating images has received a significant amount of attention due to the sharp increase in volumes of images. By considering the area of agriculture, a study [24] proposed a deep learning repetitive annotation approach for recognizing a variety of fruits and classifying the ripeness of oil palm fruit. The model was implemented on 3500 fruit images and achieved an accuracy of 98.7% for classifying oil palm fruit and 99.5% for recognizing a variety of fruits. CNNs are also used in agriculture for seed classification, despite the inherent limitations of traditional machine-learning approaches in extracting features and information from image data. Ref. [25] created a deep CNN based on MobileNetV2 with a simple architecture for seed classification. The model was applied to a seed dataset with 14 different seed classes and achieved an accuracy of 95% and 98% on testing and training, respectively. However, future research will need to compare various CNN architectures to determine the best model for solving the problem at hand.

Recent trends have shown that image collection has significantly increased, thereby activating further research in image classification and annotation. A technique based on bag of visual words (BoVW), which relies on ontology, has been widely used in this area. However, problems relating to ambiguities between image categories have posed challenges with regard to image classification and annotation. A study in [26] proposed a hierarchical max pooling (HMAX) model based on ontology to classify images of animals into their respective categories. The contribution of their model was the exploitation of semantic relationships between image categories as a way of eliminating the problem of ambiguity between image categories. The model performed well as it achieved an accuracy of 80%. However, HMAX is not a desirable technique for producing features; hence, our study has used an ensemble of CNNs for feature production.

CNNs have been widely employed to solve image classification problems, attributed to their power in extracting features and always making continuous breakthroughs in the field of image recognition. However, they suffer from a huge overhead, requiring a lot of time for the training process. To alleviate this challenge, a study [27] developed a hybrid of deep learning and random forest algorithm to solve an image classification problem. The sole purpose of the CNN is for feature extraction, and the classification process is handled by the random forest algorithm. Random forest (RF) has the advantages of fast training speed and high classification accuracy. The model was effective as it produced a low error rate of 9.18%. The model did not carry out a comparative assessment against other baseline classifiers, which is accounted for in our study.

A supervised deep-learning approach based on a stacked auto-encoder was used in [28] for the classification of forest areas. The study used unmanned aerial vehicle (UAV) datasets because they have been found to be quite useful for forest feature identification due to their relatively high spatial resolution. Through cross-validation, the model achieved an accuracy of 93%. However, one significant limitation of deep learning is that it requires more computing power than other machine learning algorithms.

3. Dataset

Given the scarcity of publicly available forest-type image datasets [29], we downloaded 35 images for each class from the internet [30,31]. Considering that the obtained image dataset was too limited for the proposed model, the geometric transformation data augmentation technique from the scikit-learn library in Python was employed to produce 65 more images for each class in the training dataset and 9 more images in the testing dataset. Hence, the resulting dataset constituted a total of 952 images, from which 800 were set aside for training and 152 were reserved for testing. Table 1 shows the corresponding forest-type image dataset distribution.

Table 1. Forest type image dataset distribution.

Training Images	Testing Images	Class
100	19	Grassland
100	19	Woodland
100	19	Orchards
100	19	Bare land
100	19	Logged forests
100	19	Degraded land
100	19	Sea
100	19	Buildings

Data augmentation is a technique that artificially increases the image dataset by creating additional modified copies of already existing data. Table 2 depicts the parameter configuration used in this study to perform data augmentation, where the first column represents the set of geometric properties that require fine-tuning, and the second column represents the set value for each geometric property. In this study, class labels were used as categorical data, and the label-encoder function from the scikit-learn library in Python was employed to convert string categorical data into numerical values. Class labels in this study represent image categories such as woodlands, shrubs, sea, orchards, logged forests, grassland, degraded land, and buildings. The class labels were transformed into distinct numerical values between 0 and 7. As presented in Table 3, the first column represents the transformed numerical values, and the second column represents the corresponding class labels. Since images were of different sizes, the resize function from scikit-learn was employed to resize all images to 226 $\times$ 226 pixels. For each category, 19 images were set aside for testing, and 100 images were reserved for training. Figure 1 shows a sample of the images used in the study.

Table 2. Data Augmentation Properties.

Property	Value
rotation_range	45
width_shift_range	0.2
height_shift_range	0.2
zoom_range	0.2
horizontal_flip	True
fill_mode	reflect

Table 3. Class labels.

Numerical Value	Class
0	Buildings
1	Degraded land
2	Grassland
3	Logged forests
4	Orchards
5	sea
6	Shrubs
7	Woodlands

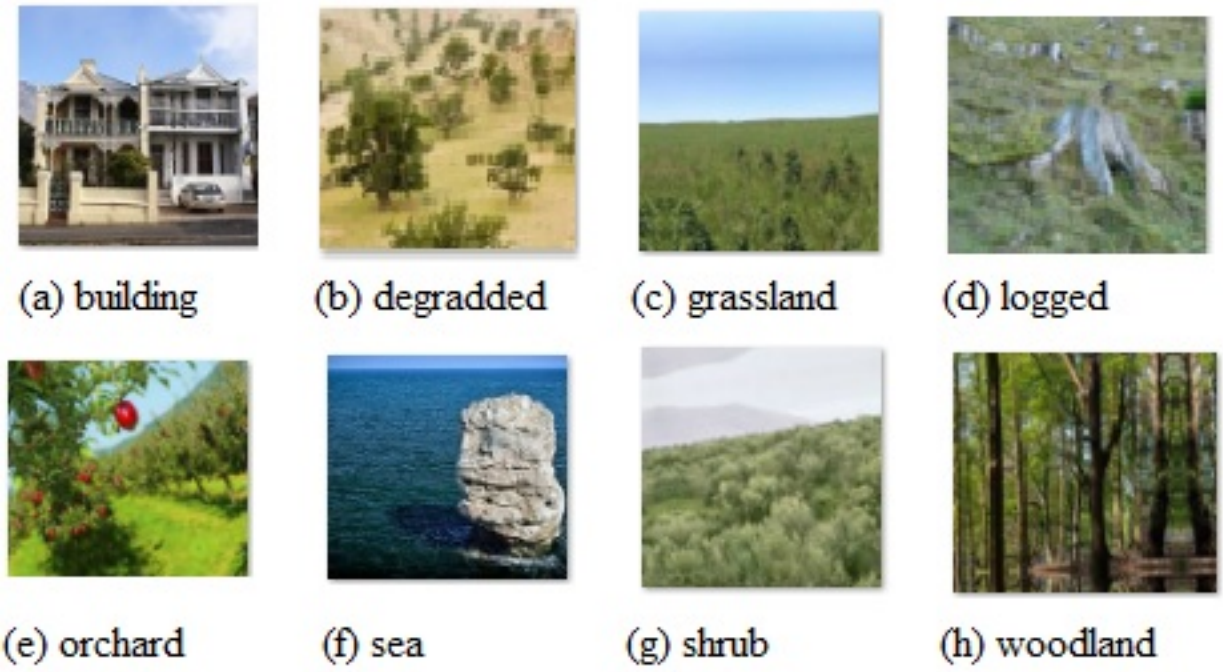


Figure 1. Sample of different types of forest images used in this study.

4. Deep Learning Architectures

4.1. Xception Architecture

Xception is expressed as "Extreme Inception". The feature extraction base of the Xception architecture has 36 convolutional layers. With the exception of the first and last modules, the convolutional layers are divided into 14 modules, all of which have linear residual connections surrounding them. The Xception architecture is briefly described as a linear stack of depthwise separable convolutional layers with residual connections. Such an architecture is very easy to define as it only takes about 30 to 40 lines of code to implement using libraries such as Keras or Tensorflow. As shown in Figure 2, images are taken as input through the entry flow section, and then subsequently channeled into the middle flow section, where the feature map process is repeated 8 times; finally, they are channeled through the exit window.

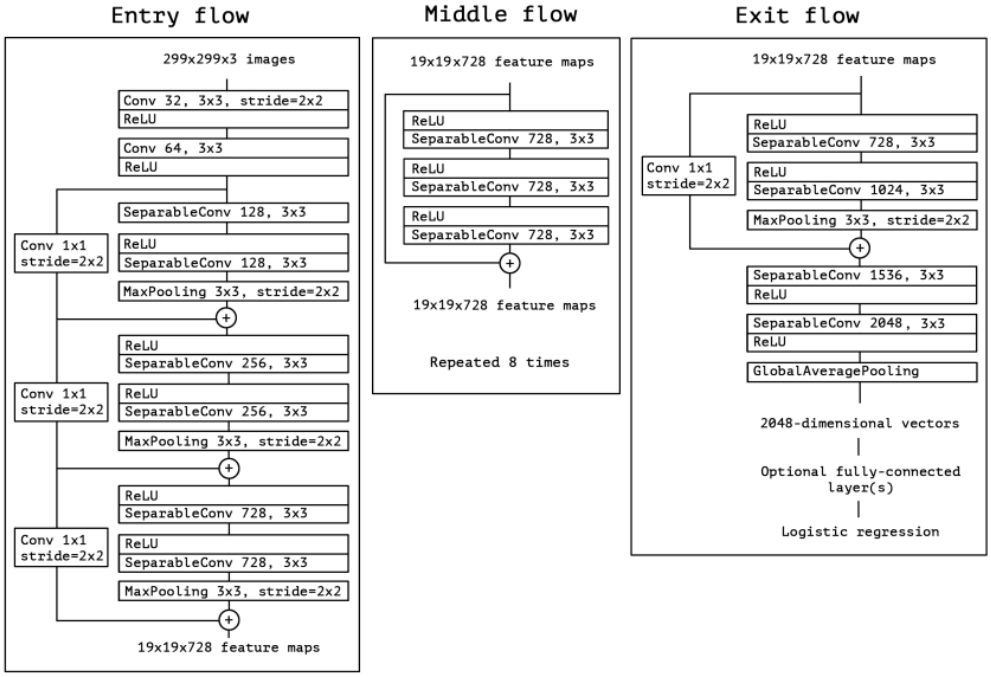


Figure 2. The Xception architecture [32].

4.2. VGG-16 Architecture

VGG-16 is a CNN network that has received huge attention in the area of computer vision due to its high classification accuracy of 92.7% when implemented on 1000 images of 1000 different categories on the ImageNet dataset [33]. The 16 in the VGG-16 architecture represents 16 convolutional layers with learnable parameters. Overall, it is composed of 13 convolutional layers, 5 pooling layers, and 3 dense layers, which gives a total of 21 layers; however, it has only 16 weight layers with learnable parameters. What is unique about VGG-16 is that it disregards a large number of hyper-parameters and instead uses 3×3 filter convolutional layers with stride 1 and max-pooling layers of 2×2 filters.

4.3. ResNet-50 Architecture

The ResNet-50 architecture was developed to overcome the degradation problem by using residual learning. It is an extremely deep type of CNN with 48 convolutional layers, 1 max pooling layer, and 1 average pooling layer. An input instance and an output instance are summed up such that the original mapping function

$$H(x) = F(x) - x \tag{1}$$

is redefined as

$$H(x) = F(x) + x. \tag{2}$$

The refinement of the mapping function greatly approximates the desired functions while also making learning simple. This reformulation was initiated to mitigate the degradation problem. The redefined mapping function in Equation (2) is implemented by having feed-forward neural networks with short connections, as shown in Figure 3.

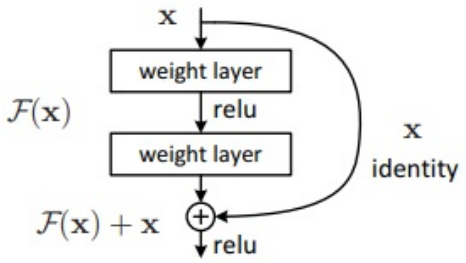


Figure 3. Building block of residual learning [34].

The shortcut connections carry out identity mapping operations, and the results are added to the outputs of the stacked layers. If the additional layers can be built as identity mappings, the training error of a deeper model should be no greater than that of its shallower counterpart.

5. Ontological Bagging Algorithm

The ontological bagging algorithm enhances semantic relationships, which in turn increases image classification accuracy. The idea behind this is to create sub-classes or categories at each ontological node based on the ontological structure. For each sub-class, weak attributes are learned, and they serve as image features for node training such that the node will be able to discriminate between the node's sub-classes.

5.1. Semantic Grouping

In order to build a hierarchical classifier, all images for all classes designated for training are required in an ontology. The naive approach of creating a semantic group involves recursively collecting only images of a particular leaf category. With the help of semantic relationships, training images for classes at the subsequent intermediate semantic levels can be accomplished by grouping together images of their offspring. At a given ontological node S, its subsequent children $S_1, \ldots, S_N$ are referred to as super-classes or categories, where N is the total number of children. Images belonging to category c_i will be denoted as m_i if c_i is a child of m_i. For instance, considering that the super-classes at the root node are "artificial crop vegetation" and "natural crop vegetation", then the training images of "natural crop vegetation" will comprise the training images of "field" and "orchard". As presented in Figure 4, logged and degraded forests are children of the "Natural Growth Vegetative Area" class rather than the "Artificial Growth Vegetative Area" class, even though both classes belong to the "Primary Vegetative Area" parent class.

5.2. Weak Attribute Learning

The bagging algorithm automatically learns many features or attributes from labeled super-category images over multiple instances. Some images belonging to one particular super-category are treated as positive bags, and the rest are treated as negative bags. Images sampled from these two bags will serve as instances of the bag. Given a super-category Q, the ontology bagging algorithm learns M unique weak features. Several image windows (r_i, t_i) are selected for each training image. Each image window r_{ij} consists of a latent variable $b_{ij} \in (0, 1, \ldots, S)$. If $b_{ij} = s \in (1, \ldots, S)$, r_{ij} denotes a positive instance of the s^{th} weak feature of S. If b_{ij} is evaluated as zero, then r_{ij} is the negative instance. Weak features are learned by giving solutions to the following objective function:

$$min_{w,bij} \sum_{s=0}^{S} ||w_s||^2 + \gamma \sum_{ij} max(0, 1 + w_{p_{ij}}^T r_{ij} - w_{b_{ij}}^T r_{ij}), \quad (3)$$

which is subject to the following constraints:

$$\begin{cases} \text{if } t_i = Q, \sum_j b_{ij} > 0 \\ \text{otherwise, } r_{ij} = 0, \end{cases} \tag{4}$$

where $p_{ij} = argmax_s \in (0, \ldots, S), s \neq b_{ij} w_s^T r_{ij}$. Each w_k represents the kth positive weak feature, while w_0 represents the negative weak feature. After learning features from all images belonging to one particular super-category at a given node, the final image features are constructed on the basis of the responses of the weak attributes or features.

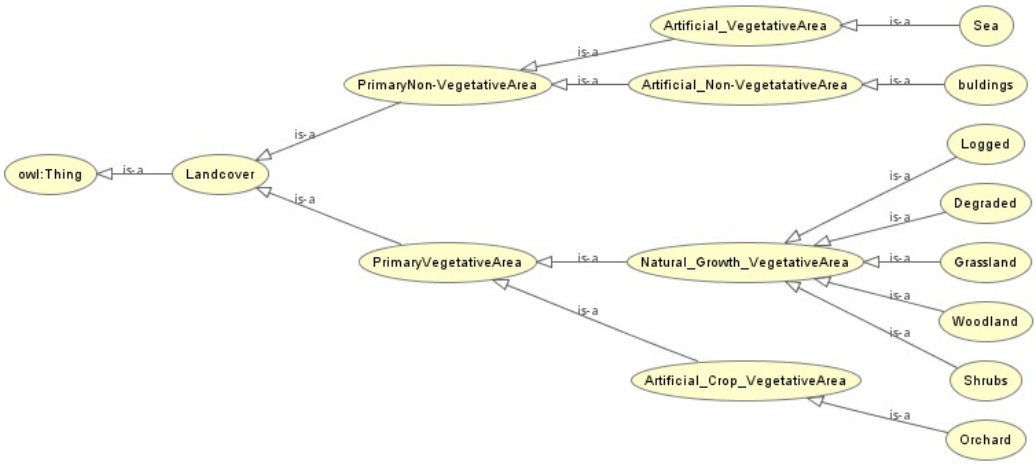


Figure 4. Ontology of Forest Types.

6. Proposed Model

In this section, the proposed image classification model and its components are explicitly described. As presented in Figure 5, the image classification approach is made up of 3 components, i.e., (1) feature extraction, (2) domain ontology construction, and (3) forest-type image classification. In the proposed model, an ensemble stack of ResNet50, Xception, and VGG16 is used to generate a feature vector (top of Figure 5) required to perform the image classification process. Image categories from the dataset are set to be used for building the ontology, which forms the basis of establishing semantic relationships between concepts of the forest domain (bottom right of Figure 5). A linear SVM multi-classifier is selected to classify images into their respective categories (bottom left of Figure 5).

6.1. Feature Extraction

The feature selection preprocessing step plays a significant role in tasks relating to image classification. The ensemble stacked model of VGG16, Xception, and ResNet50 (top of Figure 5) was used to obtain features from the training dataset. The composite set of features obtained by the three deep learning techniques was used as a feature vector for the model. The ensemble technique helps to increase the scope of the feature vector. A single feature selection method only selects an optimal subset of features from the training dataset; hence, the final feature vector may not be a true reflective set to serve as a basis for the subsequent image classification process. Ensemble feature selection may produce a more accurate outcome by combining the different outputs of various techniques.

6.2. Ontology Building

The process of building an ontology was systematically broken into two main steps: (1) concept extraction and (2) relation generation. All concepts within the forest image dataset were extracted, and relationships between these concepts were generated. In particular, this study considered only the hyponymy and hypernymy relationships. Concepts

are organized hierarchically, e.g., an image object classified as an "orchard" is an instance of the "artificial vegetation concept". The visualization of the ontology is seen as the taxonomic relationship between concepts. OWL API was used to construct the ontology with Algorithm 1.

Algorithm 1 An algorithm for ontology construction

Input : $forest_1$: classes,x: lexical resource
Output: θ: Ontology
Initialisation: $\theta \leftarrow (root : vegetation)$
concepts $\leftarrow extractconcepts(forest_1)$
subconcepts $\leftarrow findhypocon(root, concepts, x)$
while $(|subconcepts| > 0)$ **do**
    foreach $(S \in subconcepts)$ **do**
        hyperconcepts $\leftarrow findhypercon(\theta, S, x)$
        T $\leftarrow createTaxonomicR(hyperconcept, S)$
        AddTaxonomic (θ, T)
    Endforeach
    subconcepts $\leftarrow findhypocon(subconcepts, concept, x)$
end while
Return θ

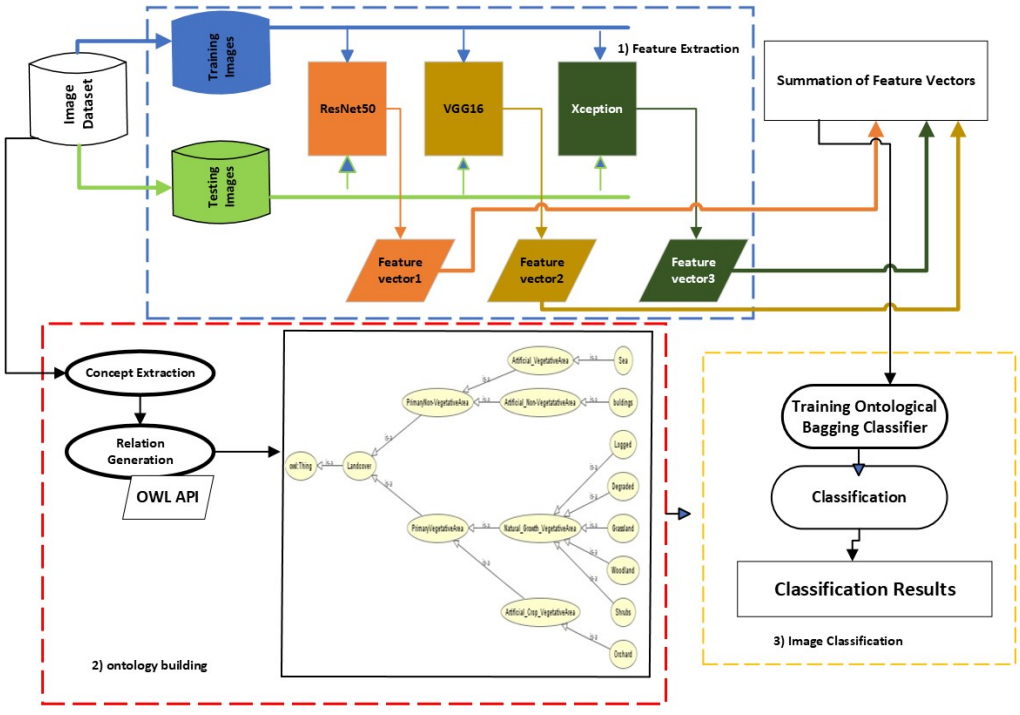


Figure 5. The proposed model.

6.3. Image Classification

This section describes the image classification process. As presented in Figure 5, the feature vector is obtained through an ensemble stacked model, and the ontology is used to perform an image classification task. The features are provided as input to a linear SVM classifier. Linear SVM is appropriate for all cases where there is a diversity of image categories [11]. The feature vector obtained from training images is used to train a one-vs.-all SVM classifier for each category in order to distinguish a given category from other categories. Each classifier will compute a confidence value that will be used to determine the appropriate category of an image. The training is based on the taxonomical relationship between ontology concepts. To begin with, all categories at each node of the ontology are

bagged into a super-category in order to train hypernym classifiers. The classification of a given test image is carried out using both the hyponymy and hypernymy classifiers, as shown in Figure 6. A test image is allocated to a category with the best hypernymy classifier, i.e., "artificial_crop vegetation" (because it has the highest confidence value). The same test image is also assigned to the best hyponymy classifier (grassland), in which the classification process is performed using the hyponymy classifiers. If the best hyponymy class and the best hypernymy class have a direct relationship, the output of both classifiers will be merged together by combining their confidence values. If there is no direct relationship between the classifiers, the best hyponymy class will be considered.

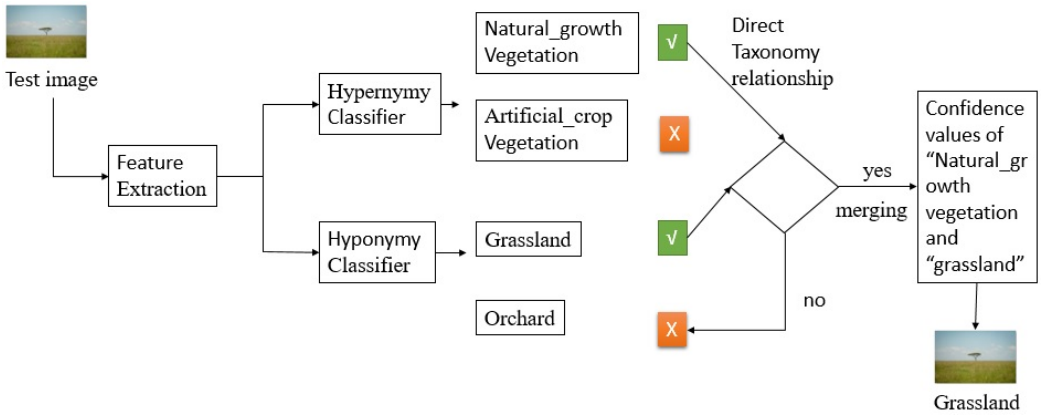


Figure 6. Classification using merging classifiers.

7. Experimental Setup

The experiments were carried out on the Google Colab platform, which offers free TPU and GPU on cloud resources. Training with GPU is faster than without GPU. Three deep learning models, namely, ResNet50, VGG16, and Xception, were adopted via transfer learning using the Python Keras library of the GPU with a Tensorflow GPU backend to perform feature extraction on images from the dataset. The hardware and software specifications for the experiments are detailed in Table 4. The features obtained from the three deep learning models were aggregated using the sum function. Owlready2, which is a module in ontology-oriented programming in Python, was used to generate the taxonomical relationships between image categories. The resulting ontology is shown in Figure 4. The set of features obtained from the sum aggregate function was used to train the classifiers according to the taxonomical relationship between image categories.

Table 4. Hardware and software specifications for the experiment.

Hardware	Software
Processor: core i5 2.2 gigahertz RAM: 32 gigabytes Graphical Processing Unit (GPU) Hard drive: 500 gigabytes NDVIDIA, 16 gigabytes RAM	Programming language: Python version 3.9 OWLReady: under the GNU LGPL licence v3 Backend: Tensorflow GPU Deep learning API: Keras GPU

Proposed Model Evaluation Metrics

The performance of the proposed model was evaluated using metrics such as accuracy, root-mean-square error (RMSE), a confusion matrix, and the receiver operating characteristic (ROC) area under the curve (AUC), commonly referred to as the ROC curve. Accuracy is a measure of how close the obtained values are to the accepted values. Accuracy is defined in Equation (5).

$$Accuracy = \frac{TN + TP}{TN + TP + FP + FN}, \quad (5)$$

where TP, TN, FN, and FN denote true positives, true negatives, false negatives, and false positives, respectively. The root-mean-square error $RMSE$ is the square root of the mean square of all errors. Because it is scale-dependent, $RMSE$ is a good measure of accuracy for comparing forecasting errors of different models or model configurations for a specific variable but not between variables. It is calculated in Equation (6).

$$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^{n} (O_i - P_i)^2}, \quad (6)$$

where O_i are the actual values and P_i are the predicted values.

The confusion matrix helps to provide a visualization of the performance of the classifiers. The visualization platform allows for easy identification of confusion between categories or classes, e.g., it is easy to identify classes that have more mislabeled data than others. The ROC curve, also known as the measure of sensitivity, is a plot of the true-positive rate versus the false-positive rate. A model with a curve that is far from the median position indicates a better classification performance. The ROC curve approximates the performance of a model across all thresholds in a plot. The bigger the area, the better the model. One of the advantages of the ROC curve is that it facilitates the comparative evaluation of results from different models without any need to balance issues related to sensitivity and specificity.

8. Experimental Results

The aim of this study is to assess the effect of ontologies in the image classification task. With that in mind, the proposed ontology-based forest-type image classification model was compared against other baseline models such as random forest (RF), K-nearest neighbor (KNN), SVM, and Gaussian naive Bayes. The features extracted using an ensemble of deep learning models were used to train the classifiers based on an ontology that describes taxonomic relationships between image classes. For a particular test image, the classification task was performed both by the hypernym and hyponym classifiers. First, the classification process began by assigning the test image to the hyponym and hypernym classifiers with the highest confidence values. The hyponym and the hypernym classifiers ran in parallel. If there was a direct relationship between hypernym and hyponym classifiers, their confidence values were merged, and the test image was assigned to the best hyponym classier. If there was no relationship between the classifiers, the next best hyponym classifier was considered, and the same process repeats.

The results presented in Table 5 show that the ontological bagging algorithm based on linear SVM outperformed other models with respect to RMSE and accuracy. The high accuracy is attributed to the ability of the model to suppress the error propagation of hierarchical classifiers.

Table 5. Accuracy and RMSE scores of the proposed model against baseline models such as kNN, GaussianNB, SVM, RF, and decision trees.

Model	RMSE	Accuracy
kNN model	1.530	0.816
GaussianNB model	1.678	0.638
SVM model	1.048	0.888
RF model	1.094	0.862
Decision tree model	2.090	0.625
Ontological bagging model	**0.532**	**0.961**

The results were further presented in terms of the confusion matrix and ROC curves.

The confusion matrix for the kNN model is illustrated in Figure 7. It is shown that the kNN model absolutely managed to correctly classify all nineteen images for class 9. Similarly, the model correctly classified seventeen test images for classes 4 and 5, but class 5

received twelve more test images from classes 0, 1, 3, 4, and 5. The kNN performed poorly in classes 3 and 5, misclassifying seven and six test images into classes 1 and 2, respectively. The associated ROC curve for the kNN model in Figure 8 produced a perfect match for classes 7 and 4 by having an ROC AUC value of 1.0. In the corresponding confusion matrix of kNN, class 4 did not receive any false-positive test images, though two test images were misclassified into class 6.

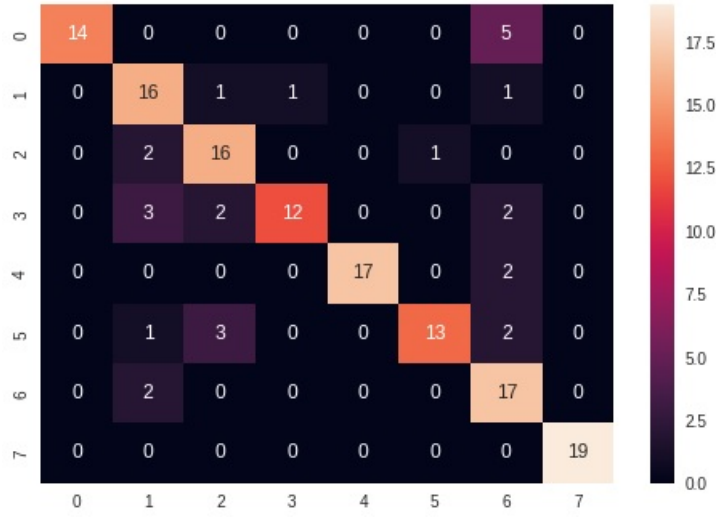


Figure 7. kNN-based confusion matrix.

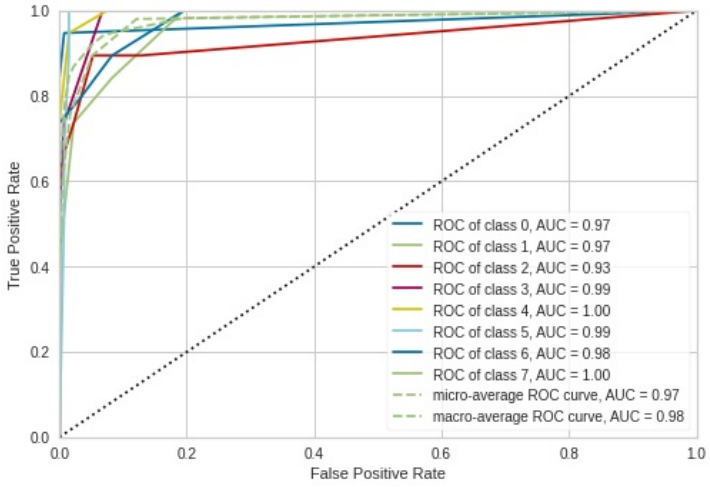


Figure 8. ROC AUC Curves for kNN model-based classifier.

The confusion matrix of our ontological bagging approach in Figure 9 provides a better alternative to image classification, as evidenced by its ability to correctly classify all nineteen test images for classes 0, 4, 5, 6, and 7, despite the fact that class 5 received two additional test images from class 2. Only one test image was misclassified for classes 1 and 3. The corresponding ROC AUC curve (Figure 10) of our model produced a perfect match for classes 0, 3, 4, 5, and 7, i.e., the model managed to precisely distinguish between positive classes and negative classes. For all the classes, the model performed the worst for class 2, and this is consistent with the corresponding results from the confusion matrix, where four false-negative test images were recorded. Class 3 obtained a perfect match because one false-positive test image and one false-negative test image canceled each other out. In contrast to the confusion matrix results, class 5 did not produce a perfect match because there was an imbalance between false positives and false negatives, as the class received more false-negative test images than false-positive test images.

As illustrated in Figure 11, the RF-based model correctly classified all nineteen test images for classes 5 and 7. Only one test image for class 0 was misclassified into class 1. The RF model performed the worst for class 2, where seven test images were misclassified into other classes. The ROC AUC curves for the RF-based classifier presented in Figure 12 produced perfect matches for classes 0, 5, and 7. These results also go in tandem with the corresponding confusion matrix results in Figure 11. All the classes have ROC AUC values that are greater than 0.9, implying that the model performed better.

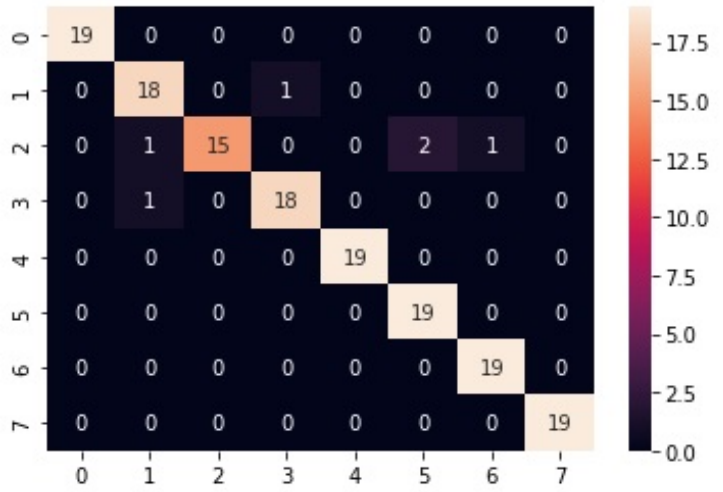


Figure 9. Ontological-bagging-based confusion matrix.

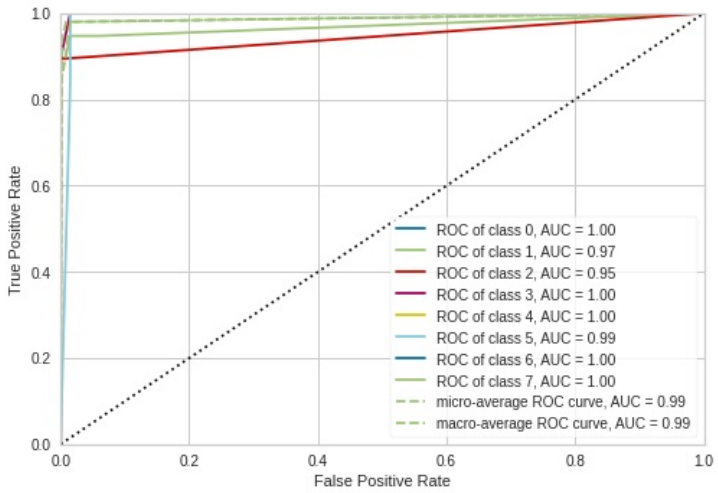


Figure 10. ROC AUC Curves for ontological-bagging-based classifier.

Overall, As shown in Figure 13, the decision-tree-based model registered the worst performance as compared to the other models for all the classes, except class 5, where all nineteen test images were correctly classified despite the fact that the same class received nine more test images from other classes. The corresponding ROC AUC curves for the decision tree in Figure 14 show that class 5 with its ROC AUC of 0.96 performed the best, and class 2 performed the worst with its ROC AUC of 0.75. This is also in line with the results obtained from the corresponding confusion matrix.

The confusion matrix of the SVM-based model presented in Figure 15 shows that a range of two to three test images out of nineteen was misclassified for classes 0, 1, 4, 5, 6, and 7. Class 2 had the worst performance with regard to the number of misclassified test images; five test images were misclassified into classes 1, 5, and 6. The associated ROC

AUC curves of the SVM-based model in Figure 16 show that class 1 with its ROC AUC value of 0.95 performs the worst, as it registered fourteen false-positive test images.

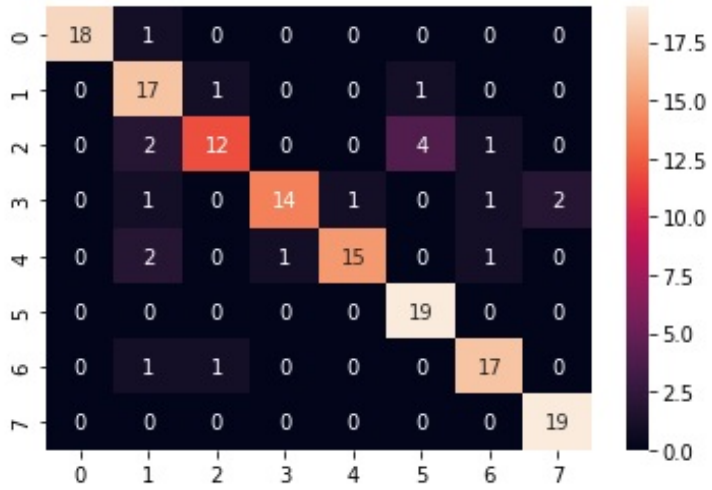


Figure 11. Random-forest-based confusion matrix.

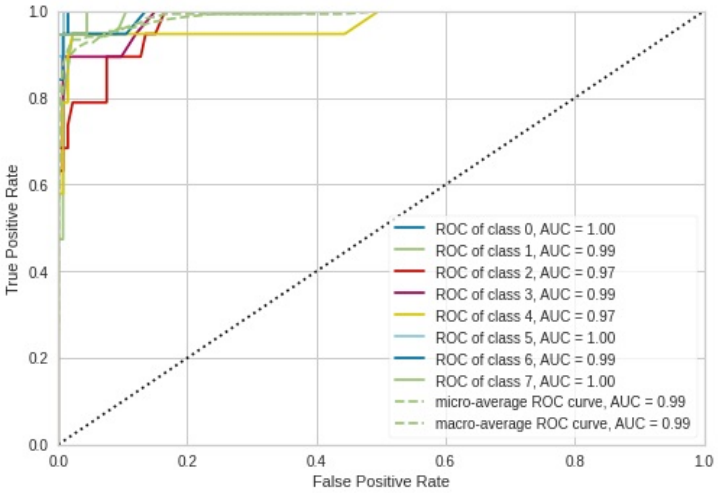


Figure 12. ROC AUC Curves for Random-Forest-based classifier.

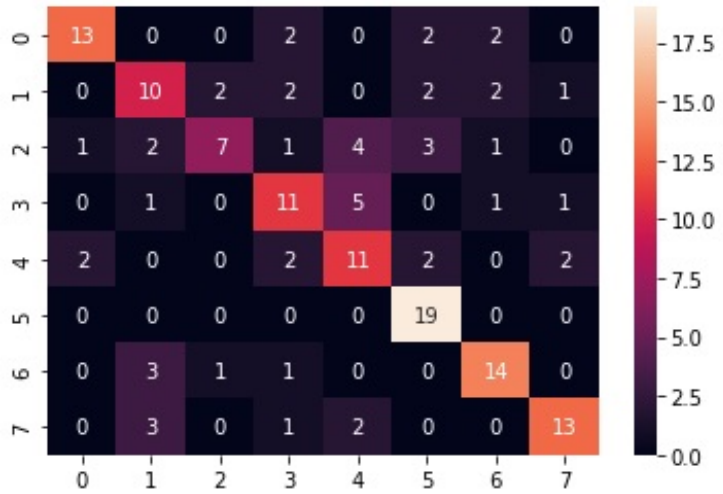


Figure 13. Decision-tree-based confusion matrix.

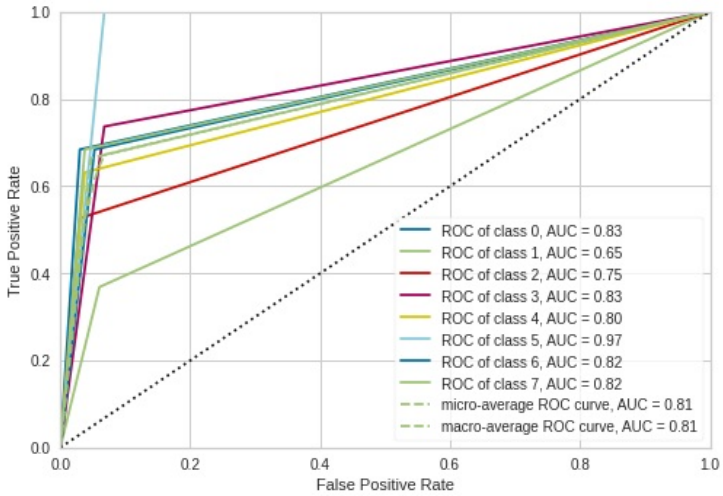


Figure 14. ROC AUC Curves for Decision-tree-based classifier.

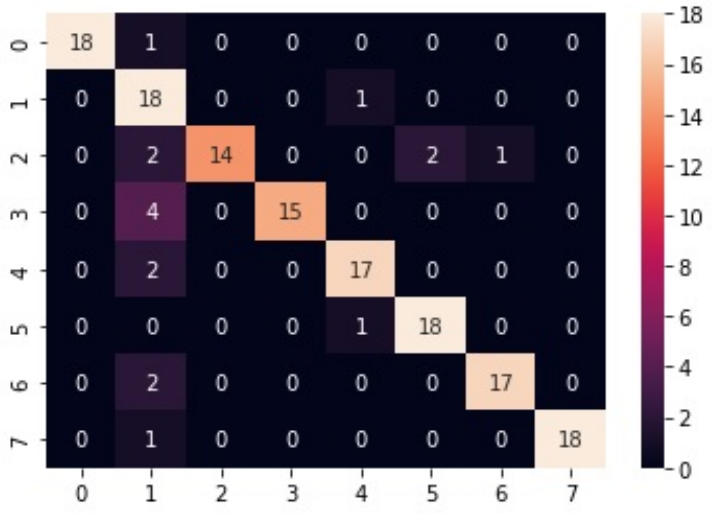


Figure 15. Support-Vector-Machine-based confusion matrix.

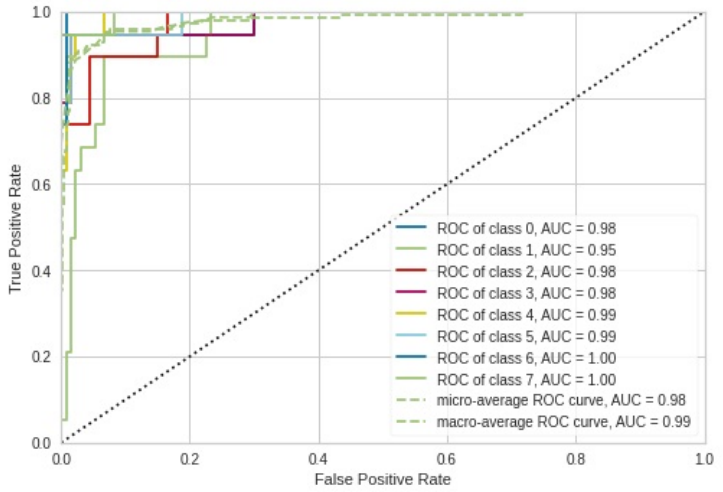


Figure 16. ROC AUC Curves for Support-Vector-Machine-based classifier.

The GaussianNB-based model presented in Figure 17 classified all nineteen test images for class 0 and class 5 despite class 5 receiving four extra test images, two from class 2

and two from class 6, and class 0 receiving 6 extra test images from classes 1, 2, 3, 4 and 7. However, the GaussianNB under-performed in classes 1 and 6, misclassifying 16 and 15 test images, respectively. Most of the test images for class 1 were misclassified into class 3, implying that most degraded land images were mistakenly viewed as logged forest images. With reference to the ROC AUC curves presented in Figure 18, the GaussianNB-based model performed best for classes 0 and 5 and performed extremely poorly for class 2. These findings also go in hand with the confusion matrix results presented in Figure 17.

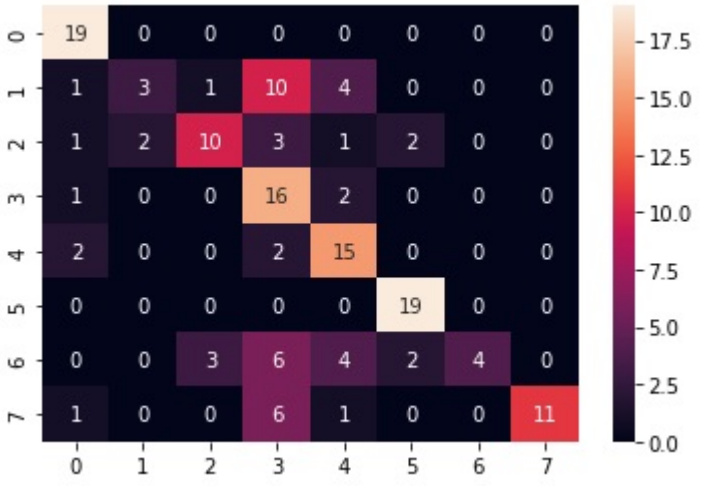


Figure 17. Gaussian-naive-Bayes-based confusion matrix.

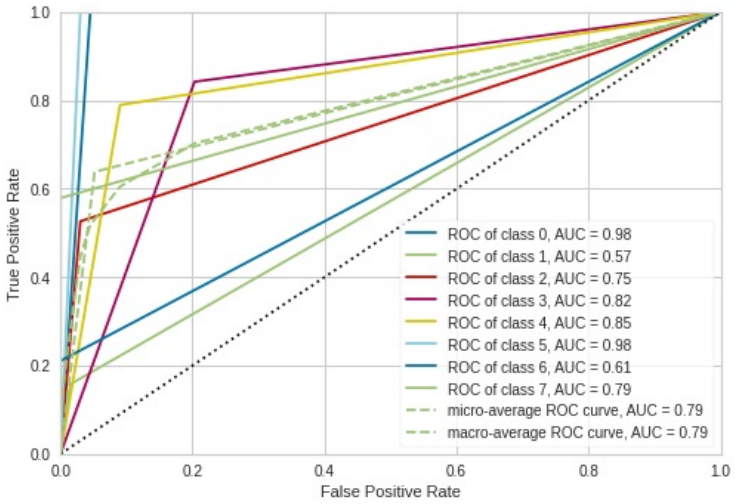


Figure 18. ROC AUC Curves for Gaussian-naive-Bayes-based classifier.

Table 6 shows that our proposed ontological bagging approach outperformed other classifiers in terms of accuracy, RMSE, and ROC_AUC. The results also demonstrate that our model has the strongest predictive power as it managed to correctly classify 146 out of 152 test images, followed by SVM, which correctly classified 135 test images. Our model registered the lowest RMSE of 0.532, implying that the model's predictions are much closer to the actual values as compared to other models. Alongside RF, our ontological bagging algorithm recorded the highest ROC_AUC value of 0.99, meaning that the model did well in separating classes as compared to other models. GaussianNB performed the worst out of all the classifiers in terms of ROC_AUC and accuracy and misclassified 55 test images into the wrong classes. The outright performance of our model is attributed to the adoption of semantic relationships between image categories for the classification process; additionally, the bagging concept helped to minimize the error propagation of classifiers.

Table 6. Quantitative comparison of models.

Model	Test Images	Correctly Classified	Misclassified	ROC_AUC	RMSE	Accuracy
kNN	152	124	28	0.97	1.530	0.816
Ontological Bagging	**152**	**146**	**6**	**0.99**	**0.532**	**0.961**
RF	152	131	21	0.99	1.094	0.862
Decision Tree	152	98	54	0.81	2.090	0.645
SVM	152	135	17	0.98	1.048	0.888
GaussianNB	152	97	55	0.79	1.678	0.638

9. Discussion

The evaluation of image classification results is of paramount importance in order to determine the best suitable model for a given application. Classification performance is dependent on the types of images used and the domain application. Images are generally categorized into remote sensing images, natural images, medical images, and synthetic images; therefore, the performance of image classification approaches varies according to the type of images used. It is possible that a particular algorithm produces good results in remote sensing images but poor results in synthetic images. For this study, image classifications based on an ontology with deep learning were obtained for natural forest images. The classes used for the study were grassland, orchards, bare land, degraded forest, woodlands, sea, buildings, and shrubs. The results presented in Table 2 show that the ontological bagging algorithm based on linear SVM outclassed other models with respect to RMSE and accuracy. The high accuracy is attributed to the ability of the model to suppress the error propagation of hierarchical classifiers. As presented in Table 7, our ontological-based model managed to outperform other models such as [11], which used ontology and an HMAX model to classify bird images into categories; ref. [1] for classifying vehicles into their respective categories; ref. [35] based on ontology and a CNN to classify natural images from an ImageNet dataset; and [36] for natural image classification through the transfer learning of images obtained from the Caltech-101 image dataset. However, the ontology-based classification model presented in [37] for classifying objects in urban and peri-urban areas slightly outperformed our model, with a classification score of 98%. A hybrid model of deep learning and SVM designed in [38] to perform image classification on the Fashion-MNIST, Cifar10, Cifar100, and Animal10 datasets also attained a classification accuracy of 99%. The reason could be attributed to the nature and quality of the image dataset generated by the data augmentation process used in the study.

Table 7. Accuracy obtained from other models.

Model	Accuracy
Ontology and Hmax model [11]	63%
Ontological random forest [1]	55%
Ontology and CNN [35]	67.27%
Deep learning model [36]	93.73%
Ontology and bag of visual words [39]	59%
Ontological-based model [37]	98%
Efficient deep learning combined with SVM [38]	99%
Ontological bagging algorithm based on linear SVM	**96%**

10. Conclusions

The proposed model for classifying images in this study uses features extracted by an ensemble deep learning technique to train classifiers, and the training is based on the taxonomic relationships between categories. Metrics such as accuracy, RMSE, confusion matrix, and ROC AUC curves were used to evaluate the model's performance.

Concepts related to image categories and the associated taxonomic relationship between them were both used to build the ontology. The ontology provided the graphical

semantic information that describes the training images. Hypernym classifiers were trained recursively using features obtained from each super-image category. Lastly, the test images were classified into their respective classes by using both the hypernym and hyponym classifiers. It is noteworthy that the proposed model of harmonizing deep learning models and ontology obtained superior performance when compared to baseline methods. The ontological bagging approach can be used in the forestry domain to classify trees according to their species and to classify vegetation into different types and categories. Ontological bagging classification can also be used to categorize fruits into their respective classes in situations such as supermarkets and factories. In the future, it is recommended to employ high-resolution networks (HRNets) as an alternative to Xception, VGG16, and Resnet50. In fact, because of their ability to convert low-resolution representation to high-level representation, which is associated with efficient block architectures developed according to new standards, they are excellent for vision tasks, such as feature extraction, semantic segmentation, and object detection [40].

Author Contributions: Introduction and related work, J.V.F.-D. Model design, M.G. Implementation and discussion section, C.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author (C. Kwenda). The forest image dataset that was generated through the data augmentation process was obtained from [30,31]. The authors confirm that the data supporting the findings of this study are available within the article.

Acknowledgments: The authors thank the University of KwaZulu Natal for providing financial assistance in accessing all resources and tools required to undertake this study.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Xu, N.; Wang, J.; Qi, G.; Huang, T.S.; Lin, W. Ontological random forests for image classification. In *Computer Vision: Concepts, Methodologies, Tools, and Applications*; IGI Global: Hershey, PA, USA, 2018; pp. 784–799.
2. Collin, C.A.; Mcmullen, P.A. Subordinate-level categorization relies on high spatial frequencies to a greater degree than basic-level categorization. *Percept. Psychophys.* **2005**, *67*, 354–364. [CrossRef] [PubMed]
3. Griffin, G.; Holub, A.; Perona, P. Caltech-256 Object Category Dataset. 2007. Available online: https://resolver.caltech.edu/CaltechAUTHORS:CNS-TR-2007-001 (accessed on 2 September 2022).
4. Fei-Fei, L.; Fergus, R.; Perona, P. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop, Washington, DC, USA, 27 June–2 July 2004; p. 178.
5. Shao, M.; Li, S.; Liu, T.; Tao, D.; Huang, T.S.; Fu, Y. Learning relative features through adaptive pooling for image classification. In Proceedings of the 2014 IEEE International Conference on Multimedia and Expo (ICME), Chengdu, China, 14–18 July 2014; pp. 1–6.
6. Griffin, G.; Holub, A.; Perona, P. Caltech-UCSD Birds 200. 2010. Available online: https://resolver.caltech.edu/CaltechAUTHORS:20111026-155425465 (accessed on 2 September 2022).
7. Biswas, S.; Chatterjee, S.; Majee, A.; Sen, S.; Schwenker, F.; Sarkar, R. Prediction of covid-19 from chest ct images using an ensemble of deep learning models. *Appl. Sci.* **2021**, *11*, 7004. [CrossRef]
8. He, T.; Zhou, H.; Xu, C.; Hu, J.; Xue, X.; Xu, L.; Lou, X.; Zeng, K.; Wang, Q. Deep Learning in Forest Tree Species Classification Using Sentinel-2 on Google Earth Engine: A Case Study of Qingyuan County. *Sustainability* **2023**, *15*, 2741. [CrossRef]
9. Ahmad, A.M.; Minallah, N.; Ahmed, N.; Ahmad, A.M.; Fazal, N. Remote sensing based vegetation classification using machine learning algorithms. In Proceedings of the 2019 International Conference on Advances in the Emerging Computing Technologies (AECT), Al Madinah Al Munawwarah, Saudi Arabia, 10 February 2020; pp. 1–6.

10. Joseph, J.L.; Kumar, V.A.; Mathew, S.P. Fruit classification using deep learning. In Innovations in Electrical and Electronic Engineering, Proceedings of the ICEEE 2021, Torino, Italy, 2–3 January 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 807–817.
11. Filali, J.; Zghal, H.B.; Martinet, J. Ontology and hmax features-based image classification using merged classifiers. In Proceedings of the International Conference on Computer Vision Theory and Applications 2019 (VISAPP'19), Prague, Czech Republic, 25–27 February 2019.
12. Filali, J.; Zghal, H.B.; Martinet, J. Comparing HMAX and BoVW Models for Large-Scale Image Classification. *Procedia Comput. Sci.* **2021**, *192*, 1141–1151. [CrossRef]
13. Guo, Y.; Gu, S. Multi-label classification using conditional dependency networks. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011.
14. Frome, A.; Corrado, G.S.; Shlens, J.; Bengio, S.; Dean, J.; Ranzato, M.; Mikolov, T. Devise: A deep visual-semantic embedding model. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–8 December 2013; Volume 26.
15. Cisse, M.M.; Usunier, N.; Artieres, T.; Gallinari, P. Robust bloom filters for large multilabel classification tasks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–8 December 2013; Volume 26.
16. Cabral, R.; Torre, F.; Costeira, J.P.; Bernardino, A. Matrix completion for multi-label image classification. In Proceedings of the Advances in Neural Information Processing Systems, Granada, Spain, 12–15 December 2011; Volume 24.
17. Griffin, G.; Perona, P. Learning and using taxonomies for fast visual categorization. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
18. Marszalek, M.; Schmid, C. Semantic hierarchies for visual object recognition. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–7.
19. Deng, J.; Ding, N.; Jia, Y.; Frome, A.; Murphy, K.; Bengio, S.; Li, Y.; Neven, H.; Adam, H. Large-scale object classification using label relation graphs. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 48–64.
20. Wu, J.; Yu, Y.; Huang, C.; Yu, K. Deep multiple instance learning for image classification and auto-annotation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3460–3469.
21. Wang, J.; Yang, Y.; Mao, J.; Huang, Z.; Huang, C.; Xu, W. Cnn-rnn: A unified framework for multi-label image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2285–2294.
22. Aggarwal, S.; Gupta, S.; Gupta, D.; Gulzar, Y.; Juneja, S.; Alwan, A.A.; Nauman, A. An Artificial Intelligence-Based Stacked Ensemble Approach for Prediction of Protein Subcellular Localization in Confocal Microscopy Images. *Sustainability* **2023**, *15*, 1695. [CrossRef]
23. Gulzar, Y. Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique. *Sustainability* **2023**, *15*, 1906. [CrossRef]
24. Mamat, N.; Othman, M.F.; Abdulghafor, R.; Alwan, A.A.; Gulzar, Y. Enhancing Image Annotation Technique of Fruit Classification Using a Deep Learning Approach. *Sustainability* **2023**, *15*, 901. [CrossRef]
25. Hamid, Y.; Wani, S.; Soomro, A.B.; Alwan, A.A.; Gulzar, Y. Smart seed classification system based on MobileNetV2 architecture. In Proceedings of the 2022 2nd International Conference on Computing and Information Technology (ICCIT), Tabuk, Saudi Arabia, 25–27 January 2022; pp. 217–222.
26. Filali, J.; Zghal, H.B.; Martinet, J. Ontology-based image classification and annotation. *Int. J. Pattern Recognit. Artif. Intell.* **2020**, *34*, 2040002. [CrossRef]
27. Xi, E. Image classification and recognition based on deep learning and random forest algorithm. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 2013181. [CrossRef]
28. Haq, M.A.; Rahaman, G.; Baral, P.; Ghosh, A. Deep learning based supervised image classification using UAV images for forest areas classification. *J. Indian Soc. Remote Sens.* **2021**, *49*, 601–606. [CrossRef]
29. Tang, Y.; Feng, H.; Chen, J.; Chen, Y. ForestResNet: A deep learning algorithm for forest image classification. *J. Phys. Conf. Ser.* **2021**, *2024*, 012053. [CrossRef]
30. Images, G. Forest. 2023. Available online: https://www.istockphoto.com/photos/forest (accessed on 2 January 2023).
31. Punnet, B. Intel Image Classification Image Scene Classification of Multiclass. 1999. Available online: https://www.kaggle.com/datasets/puneet6060/intel-image-classification?resource=download (accessed on 30 August 2022).
32. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
33. Qassim, H.; Verma, A.; Feinzimer, D. Compressed residual-VGG16 CNN model for big data places image recognition. In Proceedings of the 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 8–10 January 2018; pp. 169–175.
34. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
35. Lei, J.; Guo, Z.; Wang, Y. Weakly supervised image classification with coarse and fine labels. In Proceedings of the 2017 14th Conference on Computer and Robot Vision (CRV), Edmonton, AB, Canada, 17–19 May 2017; pp. 240–247.

36. Bansal, M.; Kumar, M.; Sachdeva, M.; Mittal, A. Transfer learning for image classification using VGG19: Caltech-101 image data set. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *14*, 3609–3620. [CrossRef] [PubMed]
37. Durand, N.; Derivaux, S.; Forestier, G.; Wemmert, C.; Gançarski, P.; Boussaid, O.; Puissant, A. Ontology-based object recognition for remote sensing image interpretation. In Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007), Patras, Greece, 29–31 October 2007; Volume 1, pp. 472–479.
38. Tan, S.; Pan, J.; Zhang, J.; Liu, Y. CASVM: An Efficient Deep Learning Image Classification Method Combined with SVM. *Appl. Sci.* **2022**, *12*, 11690. [CrossRef]
39. Abdollahpour, Z.; Samani, Z.R.; Moghaddam, M.E. Image classification using ontology based improved visual words. In Proceedings of the 2015 23rd Iranian Conference on Electrical Engineering, Tehran, Iran, 10–14 May 2015; pp. 694–698.
40. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. *Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions*; Springer International Publishing: Berlin/Heidelberg, Germany, 2021; Volume 8. [CrossRef]

4.2.2 Conclusion

This paper describes the ontological bagging approach for classifying forest images. The classification process was based on hypernym and hyponym classifiers. The approach performed well in classifying forest images as it obtained a classification accuracy of 96%. The application of the ontological bagging approach has been found to be effective in the field of forestry for the purpose of classifying trees based on their respective species, as well as categorising vegetation into various categories. The application of ontological bagging classification is also viable for the categorization of fruits into their different classes within contexts such as supermarkets and factories. In future scenarios, it is advisable to utilise high-resolution networks (HRNets) as a viable substitute for Xception, VGG16, and Resnet50. Indeed, due to their capacity to transform low-resolution depiction into high-level depiction, which is linked to proficient block architectures devised in accordance with contemporary norms, they prove to be highly proficient in visual tasks, including feature extraction, semantic segmentation, and object detection.

5 Results and Discussion

This chapter presents a comprehensive analysis of the overall results with respect to the literature survey conducted from previous studies, results with respect to segmentation models, and results with respect to state-of-the-art ontological classification-based models for classifying forest images into their respective categories.

5.1 Challenges and Proposed solutions for recent forest image classification based methods

A survey of current methods used for satellite forest image processing was conducted and challenges encountered in these methods as well as proposals for future directions that alleviate the challenges were spelled out. Challenges unearthed from the reviewed papers involve (1) Different modes of defining vegetation concepts; The definition of vegetation concept can be based on the physical, conventional, historical, or conventional model (2) Duality of forest concepts; forests can be described from the real-world perspective as characterized by high NPP values or from the image from properties as characterized by high NDVI values (3) Ambiguity and vagueness of vegetation concepts; a challenge arose in linking attribute range values to vegetation concepts is not easy, e.g. a forest has high NDVI values. The "high" is qualitative hence the classification rule becomes vague and ambiguous (4) The last issue is on the semantic gap where there is a mismatch between the interpreted data from a given situation and the data extracted based on visual information.

A knowledge framework in the form of ontologies was proposed as a means of overcoming the aforementioned challenges encountered by current methods. Ontologies eliminate the duality of the concept of vegetation by incorporating the concept of perspectivalism, in which a vegetation concept is described separately from a field of view. It offers a distinct description of vegetation entities and vegetation objects, as well as their characteristics.

Adopting probability ontologies can also be used to combat vagueness. They employ probability sets to designate concepts of interest. Attributes in the set properties are assigned probabilities, and a statistical measure of the probability value of a geographic concept is used to determine whether it is a class member.

For alleviating the sensory gap, real-world description of forest entities is correlated with matching image point descriptions of forest objects, i.e., NDVI is correlated with NPP. On the issue of combating the semantic gap, ontologies define an image object of a forest concept based on the image feature (e.g. NDVI) and its associated value ("highNDVI"). The "highNDVI" is formalized by converting symbolic

information to numeric information e.g. NDVI > 0.7. Therefore semantic gap is significantly reduced.

The other related study surveyed GEOBIA methods for forest detection and classification. In the study, the challenges with forest cover classification with Very High-resolution (VHR) images as well as solutions to each challenge were highlighted. The study also proposes the ideal state-of-the-art ontological framework that can be employed for forest image detection and classification. The challenges that were identified in the study includes: (1) domain gaps across scenes and geographical locations, (2) lack of balanced, consistent, high-quality training data, and (3) Intra-class variability and inter-class similarity for VHR data. In order to address the challenge of domain gaps across scenes and geographical locations, the study recommended the adoption of a Geometric-consistent Generative Adversarial Network (GcGAN) that eliminates any discrepancy that may arise between labeled and unlabeled images without losing their intrinsic land cover information by translating labeled feature images from the source domain to the target domain. Another approach is to adopt models through transfer learning (TL) techniques attributed to their ability to produce a generalized classifier that minimizes gaps in the feature space.

In order to deal with imbalanced training data samples, the study recommended the adoption of an impartial semi-supervised learning approach based on an extreme gradient boosting algorithm (ISS-XGB). The ISS-XGB framework integrates several semi-supervised classifiers with the purpose of addressing multi-class classification tasks. The initial step of the model involves utilizing multi-group unlabeled data to address the issue of imbalanced training samples. Subsequently, extreme gradient boosting regression is employed to replicate the target classes using positive and unlabeled samples. To address the issue of inconsistency training samples, the study recommended a novel approach for integrating many sources of data through a technique known as multi-source data fusion. This technique involves the process of re-sampling in order to harmonize the spatial resolution across the different data sources. The technique filters training samples and has the ability to offer product correction at a fine resolution. The study proposed that the utilization of techniques such as image enhancement and restoration can effectively address the issue of insufficient data quality. Various image enhancement approaches, such as Histogram equalization and Linear congruent adjustment, have been developed to increase the quality of images by effectively adjusting parameters related to contrast, brightness, and sharpness. In order to address the issue of inter-class similarity and intra-class variation, the study recommended the Venkataramanan model that autonomously selects classes for clustering and determines the best number of classes to produce. The clusters that have been obtained are regarded as distinct classes. The issue of inter-class separation is addressed by utilizing a triplet loss function, which distinguishes features among different classes.

5.2 Segmentation framework for forest images

The image segmentation process is a crucial step toward image classification. The quality of segmentation significantly affects image classification accuracy. In this research, two techniques were employed to perform the segmentation process. In the first approach, a hybrid model of CNN in the form of ResNet50 and Random Forest algorithm was developed to segment a satellite forest image into the regions of forest and non-forest areas. ResNet50 was solely adopted to generate a feature vector for the Random forest algorithm to perform the segmentation process. The model achieved a segmentation accuracy of 94%, RMSE of 0.25, and MAE of 5.92. The other technique is an extension of the first segmentation paper that employed a hybridization approach of deep neural networks, in particular, VGG16 and ResNet50, and machine learning classifiers such as RF, LDA, LSVM, kNN, and GNB. The hybrid of VGG16 and ResNet50 was formulated to generate a comprehensive set of features for the machine learning classifiers to segment an aerial forest image into forest and non-forest regions. The results obtained from the study indicate that the model based on Random Forest (RF) had superior segmentation performance, obtaining an accuracy rate of 94% and a Root Mean Square Error (RMSE) value of 0.25. By evaluating the two techniques, it can be deduced that the RF algorithm emerges as the most optimal alternative algorithm for executing tasks pertaining to image segmentation.

5.3 Classification of satellite forest images

Two different techniques were developed for the classification phase to classify satellite images into their respective categories. In the first approach, satellite forest images were classified using a composite model developed by combining the ResNet50 deep learning model and the XGboost traditional learning technique. The sole purpose of ResNet50 was to generate a set of features for the XGBoost algorithm to perform the classification process. The image categories considered in the study were shrubs, woodlands, logged forests, grassland, degraded forests, and bare lands. The XGBoost-based model obtained a classification accuracy of 77% and managed to outperform other baseline models such as Random Forest, Light Gradient Boost Machine, and a fully connected ResNet50 model which obtained a classification accuracy of 74%, 73%, and 59% respectively.

In the second classification approach, an ontological bagging approach and an ensemble technique of CNN were developed to improve forest image classification. The aim of the study was to investigate the effect of ontologies on classification accuracy. The background of the study is that most studies neglect the concept of semantic relationships between image categories and the problem that arises is that the image categories are treated as independent when in actual fact have a strong semantic

overlap. In the model, an ensemble of Xception, ResNet50, and VGG16 was used to generate a set of plausible features for the classifiers trained through ontology to perform the image classification process.The hypernym classifiers underwent recursive training, utilising features derived from each super-image category. Finally, the test photos were categorised into their appropriate groups by the utilisation of both hypernym and hyponym classifiers. The higher performance of the suggested model, which harmonises deep learning models and ontologies, is worth noting in comparison to the baseline techniques. Our ontological-based model managed to obtain a classification accuracy of 96% outperforming other baseline classifiers without ontology. By comparing the two classification techniques it is concluded that ontologies significantly increase image classification accuracy, and this is attributed to the capacity of ontology to represent domain expert knowledge and also to consider semantic relationships between image categories in the classification process. The image categories considered in the study are shrubs, woodland, grassland, buildings, sea, logged forest, orchard, and buildings.

As presented in Table 2 (A snapshot of the results obtained in paper 6) our state-of-the-art ontological-based model managed to outperform other models such as the ontological-HMAX-based model used to classify bird images into its categories; Ontological random Forest model for classifying vehicles into their respective categories; Ontology and CNN model to classify natural images from ImageNet dataset and the deep learning model for natural image classification through transfer learning of images obtained from Caltech-101 image dataset ((reference to paper 6 in the thesis). However, the ontology-based classification model proposed by Bansal et al for classifying objects in urban and peri-urban areas slightly outperformed our model, with a classification score of 98%, and the hybrid of deep learning and SVM designed by Tan et al to perform image classification on Fashion-MNIST, Cifar10, Cifar100, and Animal10 datasets also attained classification accuracy of 99% (reference to paper 6 in the thesis). The reason could be attributed to the nature and quality of the image dataset generated by the data augmentation process used in the study.

Table 2: Accuracy obtained from other models

Model	Accuracy
Ontology and Hmax model	63%
Ontologigal random Forest	55%
ontology and CNN	67.27%
Deep learning model	93.73%
Ontology and Bag of Visual Words	59%
Ontological based model	98%
Efficient Deep learning combined with SVM	99%
ontological bagging algorithm based on linear SVM	**96%**

6 Conclusion and Future Work

This chapter provides a brief recap of the preceding chapters' work, examines the results of the objectives, and previews potential strategies that might be used to improve the accuracy and precision of forest image categorization in future studies. The objectives formulated to develop the state-of-the-art framework are as follows:

- To examine the current approaches used in forest image classification.
- To evaluate contemporary methodologies suitable for forest image classification.
- To create a structured framework tailored for segmenting forest imagery.
- To develop and implement a hybrid model combining deep learning and machine learning for classifying forest images.
- To develop a comprehensive framework integrating deep learning techniques with ontology for forest image analysis.

Objectives one and two were achieved by surveying recent methods employed in forest image classification. The contribution obtained was to analyze and correlate ontologies with deep learning models and involve domain expert knowledge in expressing the semantic relationship between forest concepts.

Objective 3 was attained through two research papers, In the first paper we developed a method based on CNN and a machine-learning model for segmenting satellite forest images into the forest and non-forest regions. ResNet50 adopted under transfer learning was employed solely for producing a set of features that were fed as input to the random forest (RF) classifier to perform the segmentation task. Random forest was chosen because it performs well for segmentation tasks over a limited data set while ResNet50 performs well at extracting appropriate features suitable for segmentation. The contribution obtained from this study is that designing a hybrid approach of harmonizing deep learning techniques and machine learning techniques significantly increases segmentation accuracy than each technique used alone. In the second paper, which is an extension of the first paper, the researchers employed a combined methodology involving VGG16 and ResNet50 deep learning models to extract a comprehensive set of features. Machine learning classifiers then utilized these features to effectively segment aerial satellite images into distinct forest and non-forest regions. The integration of VGG16 and ResNet50 deep neural networks facilitates the expansion of feature representation necessary for segmenting satellite forest images.

For objective 4, a hybrid model that combined convolutional deep learning, specifically ResNet50 and traditional machine learning (XGBoost) was designed to categorize forest images into grassland, degraded forest, bare land, woodlands, shrubs, and logged forest into their respective classes. ResNet50 was chosen because it generates a comprehensive set of features required for subsequent image classification and on the other hand XGBoost was chosen because of its high efficacy in image classification. Since the study used limited forest images, the idea of the model was to adopt ResNet50 under transfer learning for generating a set of features for the XGBoost algorithm to perform image classification. The model achieved a classification accuracy of 77% hence it can be used to classify satellite or natural images into their respective categories.

In addressing objective 5, a deep learning model with ontology was used to classify forest images into their respective categories. The eight categories that were considered in the study are: sea, building, logged forest, degraded forest, grassland, woodland, shrubs, and orchards. In this study approach, a ResNet50, VGG16, and Xception ensemble was used to produce a feature vector from the training data set for subsequent image classification. The ensemble approach helps in increasing the scope of the feature vector. The process of building the ontology followed two key steps namely; (1) concept extraction and (2) relation generation. Concepts from the forest domain were generated, and the relationship between the concepts was generated. Only hypernymy and hyponymy were considered in the study. OWL API was used to build the ontology. The classification was performed by both hypernymy and hyponymy classifiers. If both the hypernymy and hyponymy obtained the highest confidence values at a particular node, then their confidence values will be merged to produce the final output image category, otherwise, the best hypernymy class will be considered. The ontological model performed well as it achieved a classification accuracy of 96%.

The main research question that guides this study is whether the integration of ontologies and deep learning approaches has a positive impact on the accuracy of forest image classification. Based on images that were used in the context of this study, the deep learning model driven by ontologies significantly increased forest image classification accuracy. This is demonstrated by the ability of our ontological-driven model (paper 6) to achieve an image classification accuracy of 96% against other baseline models without ontologies such as the kNN-based model (82%), RF-based model (86%) Decision tree-based model (65%), SVM-based model (88%), and GaussianNB based model (64%). In summary, our findings demonstrated that integrating ontologies with deep learning has a tangible and positive impact on forest image classification accuracy. These findings contribute to the growing body of knowledge in the intersection of ontologies and deep learning, and they have potential applications in various domains related to forest monitoring and analysis.

6.1 Future Work

Deep learning models are black box in nature because of the complexity of their network structure such that it is very difficult to understand how they make decisions. Therefore, domain expert knowledge may not be certain if the model gained correct knowledge, hence undermining users' confidence in deep learning models. As a way of unpacking the process by which deep learning models reach their decision, it is imperative to adopt explainable artificial intelligence (XAI) methods such as model-agnostic and visualization methods. In future studies, it is recommended to adopt high-resolution networks (HRNets) as an alternative to traditional deep learning models, because they have the ability to convert low-resolution representation to high-resolution and have efficient block structures developed according to new standards and they are excellent at being used for feature extraction. The research has not been thoroughly assessed due to the lack of a large dataset with forest images of different categories. For future studies, it is recommended to use a large dataset to assess the accuracy of the model in forest image classification.